OM of: US-09-303-518D-571 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 6:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-571
Query length: 894
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38784	+	1572.00	2905.36	3.7e-154	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74947	+	1572.00	2905.36	3.7e-154	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38782	+	1469.50	2714.43	1.6e-143	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38783	+	1469.50	2714.43	1.6e-143	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74950	+	1469.50	2714.43	1.6e-143	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74949	+	1467.50	2710.70	2.6e-143	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60652	+	1460.50	2697.67	1.4e-142	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74948	+	1400.50	2586.30	2.3e-136	28				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38781	+	591.00	1088.05	1.5e-52	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60651	+	287.50	512.73	6.7e-21	30				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY79683	+	205.00	359.64	2.4e-12	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75338	+	204.00	357.77	3.0e-12	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75337	+	202.00	354.05	4.9e-12	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75339	+	202.00	354.05	4.9e-12	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98392	+	196.00	341.64	2.1e-11	32				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37217	+	175.00	298.93	3.7e-09	45				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60656	+	168.00	290.80	1.6e-08	29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB88540	+	168.00	289.83	1.7e-08	31				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60654	+	168.00	289.83	1.7e-08	31				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60655	+	151.00	257.21	1.0e-06	34				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB65602	+	145.00	234.92	6.5e-06	95				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	+	145.00	234.22	6.6e-06	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	+	145.00	229.14	8.0e-06	16				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955	+	141.00	233.08	1.4e-05	57				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60653	+	140.00	237.93	1.4e-05	31				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34697	+	137.00	227.97	3.3e-05	46				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW25084	+	136.00	230.48	3.5e-05	31				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18663	+	134.00	224.35	6.2e-05	38				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814	+	134.00	213.68	9.2e-05	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824	+	134.00	208.72	0.0001	16				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817	-	132.50	211.14	0.0001	99				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU40508	+	132.00	221.61	9.6e-05	35				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29192	+	131.50	223.98	9.6e-05	26				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29184	+	131.50	215.67	0.0001	55				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23029	+	131.50	195.07	0.0003	36				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03914	+	130.50	213.77	0.0002	56				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817	+	130.00	206.48	0.0002	99				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827	+	130.00	201.36	0.0003	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAG33109	+	127.50	213.77	0.0003	33				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAY20852	+	126.50	217.58	0.0003	20				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827	-	126.50	194.84	0.0007					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51580	+	126.00	204.15	0.0005					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04993	+	125.00	202.11	0.0007					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29221	+	123.00	202.28	0.0009					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146	+	122.00	200.59	0.0011					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95872	+	121.50	203.69	0.0011					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY50936	+	121.00	195.51	0.0016					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB63385	+	120.50	197.67	0.0016					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG19604	+	120.00	203.04	0.0015					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998	+	119.50	197.32	0.0020					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27981	+	119.00	200.49	0.0019					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR94957	+	119.00	184.90	0.0034					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18664	+	118.50	195.51	0.0025					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94309	+	118.50	193.42	0.0027					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39201	+	118.50	191.33	0.0029					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW85665	+	118.00	196.54	0.0026					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20114	+	118.00	187.88	0.0036					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67588	+	117.50	194.00	0.0031					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04841	+	117.50	190.68	0.0036					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAG44802	+	116.50	196.69	0.0034					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG09327	+	116.00	196.26	0.0037					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU46556	+	115.50	194.31	0.0044					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU03592	+	115.50	193.49	0.0045					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY51614	+	115.50	192.48	0.0047					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG08343	+	115.50	189.59	0.0052					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14914	+	115.50	184.40	0.0063					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365	+	115.50	178.44	0.0079					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY21853	+	115.00	194.45	0.0047					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY53044	+	114.50	181.94	0.0082					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14734	+	114.00	176.24	0.0110					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU40538	+	113.50	189.08	0.0074					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504	+	113.50	175.09	0.0125					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503	+	113.50	173.63	0.0132					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74327	+	112.50	188.27	0.0091					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB22112	+	112.50	184.80	0.0103					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM40777	+	112.50	179.03	0.0128					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	-	112.50	173.69	0.0156					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	-	112.50	168.61	0.0189					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61380	+	112.00	176.74	0.0152					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95146	+	112.00	170.97	0.0188					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY95934	+	111.50	185.69	0.0118					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY05097	+	111.50	185.17	0.0121					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB87879	+	111.50	185.17	0.0121					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21540	+	111.50	178.97	0.0152					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB31659	+	111.50	177.96	0.0158					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03877	+	110.50	184.75	0.0145					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY05839	+	110.50	168.28	0.0268					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW43397	+	110.00	182.53	0.0172					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38613	+	110.00	181.81	0.0176					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG19764	+	110.00	181.81	0.0176					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18658	+	110.00	181.78	0.0177					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21245	+	110.00	173.36	0.0242					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731	+	110.00	173.20	0.0243					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU12176	+	109.50	174.01	0.0257					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY52443	+	109.00	174.69	0.0273					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG19506	+	109.00	169.84	0.0327					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU42728	+	108.50	182.87	0.0219					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM40855	+	108.50	180.22	0.0241					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY25601	+	108.50	179.37	0.0249					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG11513	+	108.50	171.33	0.0336					

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38784

seq_documentation_block:
ID AAY38784 standard; Protein; 297 AA.
XX
AC AAY38784;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigenic protein encoded by ORF138.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

```
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12219.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 328; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   297 AA;


alignment_scores:
            Quality: 1572.00              Length:    297
              Ratio:  5.293                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-303-518D-571 x AAY38784   ..

 Align seg 1/1  to: AAY38784  from: 1  to: 297

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuSerLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspThrGlnThrValLysAlaValPheAlaGluThrAlaLysCysG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheLysLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sGlyGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheHisLeuThrAlaMetTyr 150

     451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrGlyIleGlnGlyValLysGlnI 184

     551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgAlaGlyGluAlaThrIleIleLeuProAspHis 200

     601 GTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTCGGCAA 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGlyValTrpAlaAspPhePheGlyLy 217

     651 ACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 sProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysGlyV 234

     701 TGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTC 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 alLysThrLeuPhePheCysCysGluArgLeuProAspGlyGlnGlyPhe 250

     751 GTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCA 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ValLeuHisIleArgProValGlnGlyGluLeuAsnGlyAsnLysAlaHi 267

     801 CGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTC 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 sAspAlaAlaValPheAsnArgAsnThrGluTyrTrpIleArgArgPheP 284

     851 CGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
         ||||||||||||||||||||||||||||||||||||||||||
     284 roThrGlnTyrLeuPheMetTyrAsnArgTyrLysThrPro 297

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74947

seq_documentation_block:
ID   AAY74947 standard; Protein; 297 AA.
XX
AC   AAY74947;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 505 protein sequence SEQ ID NO:1368.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
```

```
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53709.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 744; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   297 AA;


alignment_scores:
            Quality: 1572.00              Length:    297
              Ratio:  5.293                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-303-518D-571 x AAY74947   ..

 Align seg 1/1  to: AAY74947  from: 1  to: 297

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuSerLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspThrGlnThrValLysAlaValPheAlaGluThrAlaLysCysG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheLysLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sGlyGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheHisLeuThrAlaMetTyr 150

     451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrGlyIleGlnGlyValLysGlnI 184

     551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgAlaGlyGluAlaThrIleIleLeuProAspHis 200

     601 GTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAA 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGlyValTrpAlaAspPhePheGlyLy 217

     651 ACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 sProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysGlyV 234

     701 TGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTC 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 alLysThrLeuPhePheCysCysGluArgLeuProAspGlyGlnGlyPhe 250

     751 GTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCA 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ValLeuHisIleArgProValGlnGlyGluLeuAsnGlyAsnLysAlaHi 267

     801 CGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTC 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 sAspAlaAlaValPheAsnArgAsnThrGluTyrTrpIleArgArgPheP 284

     851 CGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
         |||||||||||||||||||||||||||||||||||||||||
     284 roThrGlnTyrLeuPheMetTyrAsnArgTyrLysThrPro 297

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38782

seq_documentation_block:
ID   AAY38782 standard; Protein; 298 AA.
XX
AC   AAY38782;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
```

```
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12217.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 326; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1469.50              Length:    298
              Ratio:  5.085                Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 94.295

alignment_block:
 US-09-303-518D-571 x AAY38782   ..

 Align seg 1/1  to: AAY38782  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
         |||||||||||||||||||||||||||||||||||||   |||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
         |||||||   :::||||||||||||||||||||||||||||||||   |
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
         ||||||||||||||||||||||:::||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         |   |||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
         |||||||||||||||||||||||||||||||   ||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||:::|||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCCGACCAC 600
         ||||||||||||||||:::|||||||||||||||:::|||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
         ||||||||||||||||||||||   |||||||||:::||||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
         |||||||||||||||||||||||||||||||||||   ||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     748 TTCGTGTTGCACATCCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
         |||   |||||||||||||||||||||||||||||||||||:::|||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
         ||||||||||||||||||||||||||:::||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     848 TTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
         |||||||||||||||||||||||||||||||||||   |||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38783

seq_documentation_block:
ID   AAY38783 standard; Protein; 298 AA.
XX
AC   AAY38783;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
```

```
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12218.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 327; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   298 AA;


alignment_scores:
           Quality: 1469.50            Length:    298
             Ratio:  5.085              Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 94.295

alignment_block:
 US-09-303-518D-571 x AAY38783   ..

 Align seg 1/1  to: AAY38783  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
         |||||||||||||||||||||||||||||||||||||||||||   |||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
         |||||||   :::|||||||||||||||||||||||||||||||||   |
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
         ||||||||||||||||||||||||:::||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         |   |||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||   ||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProProheProLeuThrAlaMetTyr 150

     451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
         |||||||||||||||||||||||||||||:::||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
         ||||||||||||||||:::|||||||||||||||:::|||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
         |||||||||||||||||||||||   |||||||||:::|||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
         |||||||||||||||||||||||||||||||||||   |||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     748 TTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
         |||   ||||||||||||||||||||||||||||||||:::|||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
         ||||||||||||||||||||||||||:::|||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     848 TTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
         |||||||||||||||||||||||||||||||||||||   |||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74950

seq_documentation_block:
ID   AAY74950 standard; Protein; 298 AA.
XX
AC   AAY74950;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 505 protein sequence SEQ ID NO:1374.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53712.
XX
```

```
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 747; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1469.50              Length:    298
              Ratio:  5.085                Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 94.295

alignment_block:
 US-09-303-518D-571 x AAY74950   ..

 Align seg 1/1  to: AAY74950  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
         |||||||||||||||||||||||||||||||||||||||   |||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
         |||||||   :::||||||||||||||||||||||||||||||||   |
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
         ||||||||||||||||||||||||:::||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         |   |||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
         |||||||||||||||||||||||||||||||   |||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
         |||||||||||||||||||||||||||||:::|||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
         ||||||||||||||||:::||||||||||||||:::|||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
         ||||||||||||||||||||||||   ||||||||||:::|||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
         ||||||||||||||||||||||||||||||||||||   |||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     748 TTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
         |||   |||||||||||||||||||||||||||||||||:::|||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
         ||||||||||||||||||||||||||::||||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     848 TTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
         |||||||||||||||||||||||||||||||||||   |||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74949

seq_documentation_block:
ID   AAY74949 standard; Protein; 298 AA.
XX
AC   AAY74949;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 505 protein sequence SEQ ID NO:1372.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53711.
XX
```

```
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 746; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1467.50              Length:    298
              Ratio:  5.078                Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 93.960

alignment_block:
 US-09-303-518D-571 x AAY74949   ..

 Align seg 1/1  to: AAY74949  from: 1  to: 298

      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

     51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
        |||||||||||||||||||||||||||||||||||||||||   |||||||
     17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

    101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

    151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
        ||||||||||||||||||||||||||||||||||||||||||||||:::||
     51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyMetAs 67

    201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
        |||||||   :::|||||||||||||||||||||||||||||||||   |
     67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

    251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
        ||||||||||||||||||||||||:::||||||||||||||||||||||
     84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

    301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

    351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
        |   ||||||||||||||||||||||||||||||||||||||||||||||||
    117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

    401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
        ||||||||||||||||||||||||||||||   ||||||||||||||||
    134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

    451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

    501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
        ||||||||||||||||||||||||||||||:::|||||||||||||||||
    167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

    551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
        ||||||||||||||||:::|||||||||||||||:::||||||||||||
    184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

    601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
        |||||||||||||||||||||||   |||||||||:::|||||||||||
    201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

    648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

    698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
        ||||||||||||||||||||||||||||||||||||   |||||||||
    234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

    748 TTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
        |||   |||||||||||||||||||||||||||||||||:::|||||
    251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

    798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
        ||||||||||||||||||||||||||:::|||||||||||||||||||
    267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

    848 TTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
        |||||||||||||||||||||||||||||||||||   |||
    284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60652

seq_documentation_block:
ID   AAB60652 standard; Protein; 298 AA.
XX
AC   AAB60652;
XX
DT   04-MAY-2001  (first entry)
XX
DE   N. meningitidis (serogroup B) HtrB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; HtrB.
XX
OS   Neisseria meningitidis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91451.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 98; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
```

```
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following: Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Neisseria
CC   meningitidis HtrB protein.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1460.50              Length:    298
              Ratio:  5.071                Gaps:      1
 Percent Similarity:  96.644   Percent Identity: 93.624

alignment_block:
 US-09-303-518D-571 x AAB60652   ..

 Align seg 1/1  to: AAB60652  from: 1  to: 298

      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
        ||||||||||||||||   ||||||||||||||||||||||||||||||
      1 MetPheArgLeuGlnPheGlyLeuPheProProLeuArgThrAlaMetHi 17

     51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
        |||||||||||||||||||||||||||||||||||||||   |||||||
     17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

    101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

    151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
        ||||||||||||||||||||||||||||||||||||||||||||||:::||
     51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyMetAs 67

    201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
        |||||||   :::||||||||||||||||||||||||||||||||   |
     67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysG1yG 84

    251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
        |||||||||||||||||||||||||:::||||||||||||||||||||||
     84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

    301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

    351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
        |   ||||||||||||||||||||||||||||||||||||||||||||||
    117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

    401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
        ||||||||||||||||||||||||||||||||   ||||||||||||||
    134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

    451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

    501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
        ||||||||||||||||||||||||||||||:::||||||||||||||||||
    167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

    551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
        ||||||||||||||||||:::||||||||||||||||:::||||||||||||
    184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

    601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
        ||||||||||||||||||||||||   ||||||||||:::||||||||||||
    201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

    648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

    698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
        ||||||||||||||||||||||||||||||||||||   |||||||||
    234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

    748 TTCGTGTTGCACATCCGCCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
        |||   ||||||||||||||||||||||||||||||||||||:::|||||
    251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

    798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
        ||||||||||||||||||||||||||||:::||||||||||||||||||||
    267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

    848 TTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
        ||||||||||||||||||||||||||||||||||||||   |||
    284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74948

seq_documentation_block:
ID   AAY74948 standard; Protein; 288 AA.
XX
AC   AAY74948;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 505 protein sequence SEQ ID NO:1370.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
```

```
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53710.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 745; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   288 AA;


alignment_scores:
            Quality: 1400.50              Length:    286
              Ratio:  5.056                Gaps:      1
 Percent Similarity:  96.853   Percent Identity: 94.056

alignment_block:
 US-09-303-518D-571 x AAY74948   ..

 Align seg 1/1  to: AAY74948  from: 1  to: 288

      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

     51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
        |||||||||||||||||||||||||||||||||||||||||   |||||||
     17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

    101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

    151 AAGGAAGACCGCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

    201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
        |||||||   :::||||||||||||||||||||||||||||||||   |
     67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

    251 GTTTGGAACTTGCCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
        ||||||||||||||||||||||||:::||||||||||||||||||||||||
     84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

    301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

    351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
        |   ||||||||||||||||||||||||||||||||||||||||||||||||
    117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

    401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
        |||||||||||||||||||||||||||||||   ||||||||||||||||
    134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

    451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

    501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
        ||||||||||||||||||||||||||:::|||||||||||||||||||||
    167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

    551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
        ||||||||||||||||:::|||||||||||||||:::|||||||||||||
    184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

    601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
        ||||||||||||||||||||||   |||||||||:::|||||||||||||
    201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

    648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
        ||||||||||||||||||||||||||||||   ||||||||||||||||
    217 yLysProAlaTyrThrMetThrLeuAlaAla***LeuAlaHisValLysG 234

    698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
        ||||||||||||||||||||||||||||||||||||   |||||||||
    234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

    748 TTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
        |||   |||||||||||||||||||||||||||||||||||:::|||||
    251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

    798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
        ||||||||||||||||||||||||||||:::||||||||||||||||||||
    267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

    848 TTCCGACG 855
        ||||||||
    284 heProThr 286

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38781

seq_documentation_block:
ID   AAY38781 standard; Protein; 123 AA.
XX
AC   AAY38781;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
```

```
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12216.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 325; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   123 AA;


alignment_scores:
            Quality:  591.00            Length:    123
              Ratio:  5.008               Gaps:      0
 Percent Similarity:  95.935   Percent Identity: 94.309

alignment_block:
 US-09-303-518D-571 x AAY38781   ..

 Align seg 1/1  to: AAY38781  from: 1  to: 123

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
         |||||||||||||||||||||||||||||||||||||   |||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
         ||||||||||||||||||||||||||   ||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAla***MetArgGlnAlaGlyLeuAs 67

     201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
         |||||||   :::||||||||||||||||||||||||||||||||   |
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
         ||||||||||||||||||||||||:::|||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 GGGCGAAGGGCTGCTGTTC 369
         |   |||||||||||||||
     117 sHisGluGlyLeuLeuPhe 123

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60651

seq_documentation_block:
ID   AAB60651 standard; Protein; 308 AA.
XX
AC   AAB60651;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Moraxella catarrhalis HtrB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; HtrB.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91450.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 97; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Moraxella catarrhalis
CC   HtrB protein.
XX
SQ   Sequence   308 AA;


alignment_scores:
            Quality:  287.50            Length:    298
              Ratio:  1.722               Gaps:      11
 Percent Similarity:  56.040   Percent Identity: 31.544

alignment_block:
 US-09-303-518D-571 x AAB60651   ..

 Align seg 1/1  to: AAB60651  from: 1  to: 308

      64 GCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT 113
         ::::::|||      |||      :::   ||||||   ||||||   ||
```

```
  1 SerValLeuGlyPheLeuArgTyrValProLeuSerValLeuHisGlyLe 17

114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
    |:::         ::::::::::::::::::                |||
 17 uAlaAlaCysAlaSerTyrIleSerTyrHisCys.............ArgL 30

164 CGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTG.............. 198
        |||      ::::::   ||||||:::|||
 30 euSerIleTyrArgSerIle...GlnAlaAsnLeuIleLeuValHisPro 45

199 ...AACCCCGACACGCAG.........ACGGTCAAAGCCGTTTTTGCGGA 236
       ||||||:::|||            :::|||   ::::::   ::
 46 LysMetProAspAlaGlnArgGlnLysLeuAlaLysGlnIleLeuLysAs 62

237 AACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGG 286
    :          :::::::::          ::::::         |||:
 62 nGlnLeuIleSerAlaValAspSerLeuLysThrTrpAlaMetProProL 79

287 AAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAG 336
    ::         ::::::::::|||:::||||||      :::   :::
 79 ysTrpSerIleAlaGlnIleLysThrValHisHisGluAspIleLeuIle 95

337 CAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGG 386
    :::||||||   :::   :::|||:::|||   |||:::|||||||||||
 96 LysAlaLeuAlaAsnProSerGlyMetLeuAlaIleValProHisIleGl 112

387 CAGCTACGATTTGGGCGGACGCTACATCAGCCAG...CAGCTTCCGTTCC 433
    |:::::::::::   :::   :::::::::          |||
112 yThrTrpGluMetMetAsnAlaTrpLeuAsnThrPheGlySerPro.... 127

434 ACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATC 483
        |||   ||||||||||||   |||      |||:::|||::::::
128 .....ThrIleMetTyrLysProIleLysAsnAlaAlaValAspArgPhe 142

484 ATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCAT 533
    :::      ||||||   |||   :::::::::   :::||||||
143 ValLeuGlnGlyArgGluArgLeuAsnAlaSerLeuValProThrAspAl 159

534 ACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCA 583
     :::|||||||||   |||:::|||:::|||:::||||||      :::|
159 aSerGlyValLysAlaIlePheLysThrLeuLysAlaGlyGlyPheSerI 176

584 TCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGG 633
    ||||||||||||||||||||||:::|||      :::||||||
176 leIleLeuProAspHisValProAspPro......SerGlyGlyGluIle 190

634 GCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATT 683
    |||   |||||||||      :::   |||   |||||||||:::||||||
191 AlaProPhePheGlyIleLysThrLeuThrSerThrLeuAlaSerLysLe 207

684 GGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGC 733
    ||||         |||         :::   :::   |||   |||
207 uAlaAlaLysThrGlyCysAlaLeuValGlyLeuSerCysIleArgArgG 224

734 CCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTG 783
      ||||||:::|||||||   :::                  ||||||
224 luAspGlyAspGlyPheGluIlePheCysTyr............GluLeu 236

784 AACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACC..... 828
    |||               |||      ::::::::::::||||||
237 Asn..............AspGluGlnLeuTyrSerLysAsnThrLysIl 248

829 ..........................GAATATTGGATACGCCGTTTTC 850
                               |||      |||
248 eAlaThrThrAlaLeuAsnGlyAlaMetGluGlnMetIleTyrProHisP 265

851 CGACGCAGTATCTGTTTATGTACAACCGCTATAAA...ACGCCG 891
        :::|||::::::   |||:::|||:::|||   ||||||
265 heLeuHisTyrMetTrpSerTyrArgArgPheLysHisThrPro 279
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY79683

seq_documentation_block:

```
ID   AAY79683 standard; Protein; 293 AA.
XX
AC   AAY79683;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Neisseria meningitidis HtrB1 acyltransferase.
XX
KW   Lipopolysaccharide; vaccine; adjuvant; htrB1 gene;
KW   acyltransferase; toxicity; attenuation.
XX
OS   Neisseria meningitidis.
XX
PN   WO200026384-A1.
XX
PD   11-MAY-2000.
XX
PF   03-NOV-1998;   98WO-NL00633.
XX
PR   03-NOV-1998;   98WO-NL00633.
XX
PA   (NEWE-) NEDERLANDEN MIN WELZIJN.
XX
PI   Van Der Ley PA,  Hamstra HJ,  Steeghs LJJM;
XX
DR   WPI; 2000-422514/36.
DR   N-PSDB; AAA27756.
XX
PT   New recombinant lipopolysaccharide, useful as low-toxicity adjuvant for
PT   vaccines, has altered pattern of acylation and/or phosphate residues
PT   attached to glucosamine  -
XX
PS   Disclosure; Fig 2B; 40pp; English.
XX
CC   The present sequence is that of an acyltransferase encoded by the
CC   htrB1 gene (see AAA27756) of Neisseria meningitidis.  The
CC   acyltransferase is involved in the secondary acylation of
CC   lipopolysaccharide (LPS).  Mutations in the htrB1 gene provide
CC   an LPS product that is less toxic than native LPS, but has higher
CC   adjuvant activity.  The invention is directed at novel less toxic
CC   forms of LPS that are obtained through genetically modified
CC   Gram-negative bacteria.  The novel LPS has fewer secondary acyl
CC   chains per molecule of LPS than the native LPS, the secondary
CC   acyl chains being bound to primary acyl chains, and the primary
CC   acyl chains being bound to the glucosamine of the LPS molecule.
CC   Recombinant LPS is produced by cultivation of a Gram-negative
CC   bacterium, such as N. meningitidis, having a mutation in a gene
CC   encoding a protein involved in lipid A biosynthesis, particularly
CC   at the level of secondary acyl addition, especially the htrB1 gene.
CC   It is used as an adjuvant in vaccines used to stimulate an immune
CC   response against Gram-negative bacteria, particularly for
CC   controlling infections caused by organisms from which the LPS is
CC   derived, or by other organisms.  The acylation pattern of the
CC   recombinant LPS is homogeneous, which facilitates standardization.
XX
SQ   Sequence   293 AA;
```

```
alignment_scores:
           Quality:  205.00              Length:    288
             Ratio:  1.220                 Gaps:     10
 Percent Similarity:  58.333   Percent Identity: 26.042

alignment_block:
 US-09-303-518D-571 x AAY79683   ..

 Align seg 1/1  to: AAY79683  from: 1  to: 293

     67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
        :::|||      |||::::::|||   ::::::   ||||||   :::::
     11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 27

    117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
```

```
      ::::      |||   ||||||:::   |||:::|||      |||
   27 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgT 44

  167 GCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACG 216
                ||:::   ::::::      :::   :::   :::
   44 hrGlyGluIleAsnLeuAlaLysCysPheSerGluTrpSerGluGluLys 60

  217 GTCAAAGCCGTT............TTTGCGGAAACGGCAAAATGCGGTTT 254
         |||:::|||           |||   :::   ||||||      ||
   61 ArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 77

  255 GGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGT 304
      ||||   :::      ::::::                :::::::::::
   77 uGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerLeuV 94

  305 TCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGC 354
        :::      ::::::   ::::::::::::::::||||||      |||
   94 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 109

  355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
      |||   ::::::::::::   ||||||:::   ::::::::::::::
  110 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 126

  405 ACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGC 454
            ::::::|||::::::|||      |||   :::|||||||:::
  126 lTyrAlaLeuAsnGlnAspIlePro......LeuIleSerMetTyrSerH 141

  455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGC 504
          |||   |||   :::|||:::   :::      ||||||   |||
  141 isGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsnArg 157

  505 GGCAAA..............GGCAAAACCGCGCCCACCGGCATACAAGG 539
                         |||:::|||               :::||
  158 TyrHisAsnValPheLeuIleGlyArgThr..............GluGl 169

  540 GGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATC. 588
      |:::::   ::::::|||   :::|||   ::::::|||   ::::::
  169 yLeuArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheLeuT 186

  589 ..CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC.........GGC 627
        |||||||||               |||:::   |||        :::
  186 yrLeuProAsp..............GlnAspPheGlyArgAsnAspSer 197

  628 GTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGC 677
      |||::::::|||||||||||||      :::   |||:::|||     :::
  198 ValPheValAspPhePheGlyIleGlnThrAlaThrIleThrGlyLeuSe 214

  678 AAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAAC 727
      ::::::::|||   :::   ::::::|||::::::              |
  214 rArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProValA 231

  728 GCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGG 777
      ||      |||:::         :::|||||||:::   |||:::
  231 rgGluAlaAspAsnThr...ValThrLeuHisPheTyrProAlaTrpLys 246

  778 GAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCGCAA 824
      ::::::   |||:::   |||      ||||||      :::||||||
  247 SerPheProGlyGluAspAlaLysAlaAspAlaGlnArgMetAsnArgPh 263

  825 TACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACA 874
          |||      :::|||:::   |||   ||||||:::::::::::::
  263 eIleGluAspArgValArgGluHisProGluGlnTyrPheTrpLeuHisL 280

  875 ACCGCTATAAAACG 888
      ::|||:::||||||
  280 ysArgPheLysThr 284
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75338

seq_documentation_block:

```
ID   AAY75338 standard; Protein; 293 AA.
XX
AC   AAY75338;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 663 protein sequence SEQ ID NO:2150.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54100.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1054; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   293 AA;
```

alignment_scores:

```
            Quality:  204.00              Length:    289
              Ratio:  1.207                 Gaps:     12
 Percent Similarity:  58.478   Percent Identity: 27.336
```

alignment_block:

```
 US-09-303-518D-571 x AAY75338   ..

 Align seg 1/1  to: AAY75338  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
         :::|||      |||::::::|||   ::::::   ||||||   :::::
      11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 27

     117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
```

```
     ::::      |||   ||||||   |||||||||   :::   |||   |
  27 aAspLeuThrGlyLeuLeuAla...TyrLeuLeuValLysProArgArgA 43

 167 GCATC............GTCGCCAATATGCGGCAGGCGGGTTTGAACCCC 204
     |||||            :::|||:::      :::         :::
  43 rgIleGlyGluIleAsnLeuAlaLysCysPheSerGluTrpSerGluGlu 59

 205 GACACGCAGACGGTC...AAAGCCGTTTTTGCGGAAACGGCAAAATGCGG 251
           :::||||||   |||      |||   :::   ||||||
  60 LysArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMe 76

 252 TTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAA 301
      ||||||   :::      ::::::                ::::::::::
  76 tLeuGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerL 93

 302 TGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAG 351
     ::   :::      ::::::   ::::::::::::::::||||||
  93 euValArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAla 108

 352 GGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
     ||||||   ::::::::::::   ||||||:::   ::::::::::::::
 109 GlyGluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAl 125

 402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACA 451
     :       ::::::|||::::::|||      |||   :::||||||:
 125 aValTyrAlaLeuAsnGlnAspIlePro......LeuIleSerMetTyrS 140

 452 AGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTG 501
     ::      |||   |||   :::|||:::   :::      ||||||
 140 erHisGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsn 156

 502 CGCGGCAAA..............GGCAAAACCGCGCCCACCGGCATACA 536
     |||                    |||:::|||               ::
 157 ArgTyrHisAsnValPheLeuIleGlyArgThr..............Gl 168

 537 AGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCA 586
     :|||:::::::   ::::::|||   :::|||   ::::::|||   ::::
 168 uGlyLeuArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheL 185

 587 TC...CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC......... 624
     ::   |||||||||               |||:::   |||
 185 euTyrLeuProAsp..............GlnAspPheGlyArgAsnAsp 196

 625 GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGC 674
     :::|||:::::::||||||||||||      :::   |||:::|||
 197 SerValPheValAspPhePheGlyIleGlnThrAlaThrIleThrGlyLe 213

 675 GGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCG 724
      ::::::::::|||   :::   :::::::|||::::::
 213 uSerArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProV 230

 725 AACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAA 774
       |||      |||:::         :::|||||:::   |||:::
 230 alArgGluAlaAspAsnThr...ValThrLeuHisPheTyrProAlaTrp 245

 775 GGGGAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCG 821
       ::::::   |||:::   |||      ||||||      :::|||||
 246 LysSerProGlyGluAspAlaLysAlaAspAlaGlnArgMetAsnAr 262

 822 CAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGT 871
     |      |||      :::|||:::   |||   ||||||::::::::::
 262 gPheIleGluAspArgValArgGluHisProGluGlnTyrPheTrpLeuH 279

 872 ACAACCGCTATAAAACG 888
     :::::|||:::||||||
 279 isLysArgPheLysThr 284
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75337

seq_documentation_block:

```
ID   AAY75337 standard; Protein; 293 AA.
XX
AC   AAY75337;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 663 protein sequence SEQ ID NO:2148.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54099.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1053; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   293 AA;
```

```
alignment_scores:
            Quality:  202.00              Length:    288
              Ratio:  1.174                 Gaps:     10
 Percent Similarity:  59.722    Percent Identity: 25.694

alignment_block:
 US-09-303-518D-571 x AAY75337   ..

 Align seg 1/1  to: AAY75337  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
         :::|||      |||:::::::|||   ::::::   ||||||   :::::
      11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 27

     117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
```

```
      :::: :::||| ||||||::: |||::::||| |||
 27 aGlyLeuIleGlySerLeuAlaTyrLeuLeuValLysProArgArgArgI 44

167 GCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACG 216
          |||::: :::::: ::: ||| :::
 44 leGlyGluIleAsnLeuAlaLysCysPheProGluTrpAspGluGluLys 60

217 GTCAAAGCCGTT...........TTTGCGGAAACGGCAAAATGCGGTTT 254
      |||:::||| ||| ::: |||||| ||
 61 ArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 77

255 GGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGT 304
    |||| ::: ::::::: ::: ::: ::::::::::::
 77 uGluTyrGlyLeuTyrTrpTyrAlaSerAlaLysCysLeuLysSerLeuV 94

305 TCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGC 354
     ::: :::::: ::::::::::::::::|||||| |||
 94 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 109

355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
    ||| :::::::::::: ||||||::: ::::::::::::::
110 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 126

405 ACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGC 454
        ::::::|||::::::||| ||| :::|||||:::
126 lTyrAlaLeuAsnGlnAspValPro......LeuIleSerMetTyrSerH 141

455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGC 504
      ||| ||| :::|||::: ::: |||||| |||
141 isGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsnArg 157

505 GGCAAA.............GGCAAAACCGCGCCCACCGGCATACAAGG 539
                  |||:::||| :::||
158 TyrHisAsnValPheLeuIleGlyArgThr.............GluGl 169

540 GGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATC. 588
    |:::::: ::::::||| :::||| :::::::||| ::::::
169 yLeuArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheLeuT 186

589 ..CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC.........GGC 627
      ||||||||| |||::: ||| :::
186 yrLeuProAsp.............GlnAspPheGlyArgAsnAsnSer 197

628 GTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGC 677
    |||::::::|||||||||||| ::: |||:::||| ::
198 ValPheValAspPhePheGlyIleGlnThrAlaThrIleThrGlyLeuSe 214

678 AAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAAC 727
    ::::::::||| ::: ::::::|||::::::: |
214 rArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProValA 231

728 GCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGG 777
    || |||::: :::|||:::::: |||:::
231 rgGluAlaAspAsnThr...ValThrLeuGlnPheTyrProAlaTrpLys 246

778 GAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCGCAA 824
    :::::: :::::: |||::: |||||| :::||||||
247 SerPheProSerGluAspAlaGlnAlaAspAlaGlnArgMetAsnArgPh 263

825 TACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACA 874
      ||| :::|||::: ||| ||||||:::::::::::::
263 eIleGluGluArgValArgGluHisProGluGlnTyrPheTrpLeuHisL 280

875 ACCGCTATAAAACG 888
    ::|||:::|||||| 
280 ysArgPheLysThr 284

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75339

seq_documentation_block:
ID   AAY75339 standard; Protein; 293 AA.
XX
AC   AAY75339;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 663 protein sequence SEQ ID NO:2152.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54101.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 1055; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   293 AA;

alignment_scores:
            Quality:  202.00          Length:    288
              Ratio:  1.195               Gaps:     10
 Percent Similarity:  58.681   Percent Identity: 26.389

alignment_block:
 US-09-303-518D-571 x AAY75339   ..

 Align seg 1/1  to: AAY75339  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
         :::||| |||:::::|||  :::::: ||||||  |||::
      11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysLeuAl 27

     117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
```

```
     ::::      |||   ||||||:::   |||:::|||      |||
  27 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgI 44

 167 GCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACG 216
                ||:::   ::::::      :::   |||   :::
  44 leGlyGluIleAsnLeuAlaLysCysPheProGluTrpAspGlyLysLys 60

 217 GTCAAAGCCGTT...........TTTGCGGAAACGGCAAAATGCGGTTT 254
       |||:::|||            |||   :::   ||||||      ||
  61 ArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 77

 255 GGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGT 304
     ||||   :::      ::::::                :::::::::::
  77 uGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerLeuV 94

 305 TCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGC 354
       :::      ::::::   ::::::::::::::|||||      |||
  94 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 109

 355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
     |||   ::::::::::::   ||||||:::   ::::::::::::::
 110 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 126

 405 ACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGC 454
           ::::::|||::::::|||      |||   :::||||||:::
 126 lTyrAlaLeuAsnGlnAspValPro......LeuIleSerMetTyrSerH 141

 455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGC 504
       |||   |||   :::|||:::   :::      ||||||   |||
 141 isGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsnArg 157

 505 GGCAAA..............GGCAAAACCGCGCCCACCGGCATACAAGG 539
                        |||:::|||              :::||
 158 TyrHisAsnValPheLeuIleGlyArgThr..............GluGl 169

 540 GGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATC. 588
     |:::::::   ::::::|||   :::|||   :::::::|||   ::::::
 169 yLeuArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheLeuT 186

 589 ..CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC.........GGC 627
       |||||||||               |||:::   |||          :::
 186 yrLeuProAsp..............GlnAspPheGlyArgAsnAspSer 197

 628 GTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGC 677
     |||::::::|||||||||||||      :::   |||:::|||      ::
 198 ValPheValAspPhePheGlyIleArgThrAlaThrIleThrGlyLeuSe 214

 678 AAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAAC 727
     :::::::|||   :::   ::::::|||::::::                |
 214 rArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProValA 231

 728 GCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGG 777
     ||      |||:::         :::|||||||:::   |||:::
 231 rgGluAlaAspAsnThr...ValThrLeuHisPheTyrProAlaTrpGlu 246

 778 GAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCGCAA 824
     ::::::   ::::::   |||:::   ||||||      :::||||||
 247 SerPheProSerGluAspAlaGlnAlaAspAlaGlnArgMetAsnArgPh 263

 825 TACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACA 874
       |||      :::|||:::   |||   ||||||:::::::::::::
 263 eIleGluGluArgValArgGluHisProGluGlnTyrPheTrpLeuHisL 280

 875 ACCGCTATAAAACG 888
     ::|||:::||||||
 280 ysArgPheLysThr 284

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98392

seq_documentation_block:
ID   AAG98392 standard; Protein; 328 AA.
XX
AC   AAG98392;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Escherichia coli protein sequence SEQ ID NO:440.
XX
KW   Escherichia coli; identification; proliferation; microorganism;
KW   antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW   bacterial growth inhibition.
XX
OS   Escherichia coli.
XX
PN   WO200148209-A2.
XX
PD   05-JUL-2001.
XX
PF   19-DEC-2000; 2000WO-US34419.
XX
PR   23-DEC-1999;   99US-0173005.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen KL,  Zyskind JW;
XX
DR   WPI; 2001-457376/49.
DR   N-PSDB; AAH81448.
XX
PT   Novel nucleic acids encoding proteins required for Escherichia coli
PT   proliferation, useful for screening for antimicrobial agents -
XX
PS   Claim 19; Page 559; 596pp; English.
XX
CC   The present invention describes a purified or isolated nucleic acid
CC   sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC   given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC   microorganism is capable of inhibiting proliferation of a microorganism.
CC   (I) have antibacterial and antibiotic activities, and can be used in
CC   gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC   of the microorganism, and the manufactured antibiotic is useful for
CC   reducing the activity or level of a gene product required for
CC   proliferation of a microorganism in a subject, specifically humans. The
CC   nucleic acids that inhibit bacterial growth or proliferation can be used
CC   as antisense therapeutics for killing bacteria. In addition to
CC   therapeutic applications, the nucleic acid sequences complementary to
CC   sequences required for proliferation can be used as diagnostic tools.
CC   For example, nucleic acid probes complementary to proliferation-required
CC   sequences that are specific for particular species of microorganisms can
CC   be used as probes to identify particular microorganism species in
CC   clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC   proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC   represent oligonucleotides, which are used in the exemplification of the
CC   present invention.
XX
SQ   Sequence   328 AA;


alignment_scores:
           Quality:  196.00              Length:    287
             Ratio:  1.195                 Gaps:     12
 Percent Similarity:  57.143   Percent Identity: 27.526

alignment_block:
 US-09-303-518D-571 x AAG98392   ..

 Align seg 1/1  to: AAG98392  from: 1  to: 328

      72 CAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACC 121
         |||:::|||   |||::::::::::||||||::::::::   |||
      53 GlnLeuProTyrProValLeuCysPheLeuGlyThrArgIleGlyAlaMe 69

     122 GGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATC 171
           ||||||            ||||||      |||   ::::::
      69 tAlaArg..............ProPheLeuLysArgArgGluSerIleA 81
```

```
 172 GTCGC.....CAATATGCGGCAGGCGGGTTT 197
 81 laArgLysAsnLeuGluLeuCysPheProGlnHisSerAla..... 94

 198 GAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAT 247
 95 GluGluArgGlu.LysMetIleAlaGluAsnPheArgSerLeuGlyMetA 111

 248 GCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAA 297
 111 laLeuValGluThrGlyMetAlaTrpPheTrpProAspSerArgValArg 127

 298 ACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGA 347
 128 LysTrpPheAsp...ValGluGlyLeuAspAsnLeuLysArgAlaGlnMe 143

 348 CAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATT 397
 143 tGlnAsnArgGlyValMetValValGlyValHisPheMetSerLeuGluL 160

 398 TGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATG 447
 160 euGlyGlyArgValMetGlyLeuCysGlnPro......MetMetAlaThr 174

 448 TACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAG 497
 175 TyrArgProHisAsnAsnGlnLeuMetGluTrpValGlnThrArgGlyAr 191

 498 GGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAAC 547
 191 gMetArg......SerAsnLysAlaMetIleGlyArgAsnAsnLeuArgG 206

 548 AAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGAC 597
 206 lyIleValGlyAlaLeuLysLysGlyGluAlaValTrpPheAlaProAsp 222

 598 CACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGG 647
 223 Gln......AspTyrGlyArgLysGlySerSerPheAlaProPhePheAl 237

 648 C...AAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCA 694
 237 aValGluAsnValAlaThrThrAsnGlyThrTyrValLeuSerArgLeuS 254

 695 AAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAA 744
 254 erGlyAlaAlaMetLeuThrValThrMetValArgLysAlaAspTyrSer 270

 745 GGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGC..... 789
 271 GlyTyrArgLeuPheIleThrPro.....GluMetGluGlyTyrPr 284

 790 .....AACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACCG 829
 284 oThrAspGluAsnGlnAla......AlaAlaTyrMetAsnLysIleIleG 299

 830 AATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGC 879
 299 luLysGluIleMetArgAlaProGluGlnTyrLeuTrpIleHisArgArg 315

 880 TATAAAACG 888
 316 PheLysThr 318
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37217

seq_documentation_block:
ID AAY37217 standard; Protein; 455 AA.
XX
AC AAY37217;
XX
DT 07-OCT-1999 (first entry)
XX
DE Amino acid sequence of a Chlamydia trachomatis protein.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Disclosure; Page 981-982; 1755pp; English.
XX
CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SQ Sequence 455 AA;

alignment_scores:
Quality: 175.00 Length: 341
Ratio: 1.054 Gaps: 12
Percent Similarity: 48.680 Percent Identity: 24.633

alignment_block:
US-09-303-518D-571 x AAY37217 ..

Align seg 1/1 to: AAY37217 from: 1 to: 455

```
  22 CTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTG.. 69
   2 LeuPheLysArgLeuArgThrGlyGlyLysIleLeuValAspHisLeuVa 18

  70 .....CTCAAATGCCTCTCCCTGCTGTCGCTTT 97
  18 lTyrGlyLeuGlyLeuGlyValLeuThrIleLeuArgLeuLeuProArgS 35

  98 CCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTT 147
  35 erSerLeuArgLeuPheSerLysGlyLeuGlyThrAlaLeuPheTyrPhe 51

 148 TTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATG..... 183
  52 IleSerAspPheArgLysThrAlaLeuThrAsnLeuAlaLeuAlaPhePr 68

 184 .....CGGCAGGCGGGTTTGAACCCCGACACGCAGA 214
  68 oGluLysSerPheAlaGluArgTyrGlnIleAlaArgGlnSerValGlnG 85
```

```
 215 CGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA..................... 246
       :::      :::|||:::|||   |||
  85 lnMetIleIleThrPheValGluLeuAlaThrValAspLysPheAlaLys 101

 247 .......................TGCGGTTTGGAACTTGCCCCC...GC 269
                            :::     |||   ||||||   ::
 102 HisIleAspGluMetIleAlaIleAlaThrSerGluAspAlaProGluGl 118

 270 GTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACG 319
     :||||||      ||||||:::::::::::::::::   :::
 118 yPhePhe......ProGluGluValSerSerGlnGlnGluLeu....... 130

 320 GCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTC 369
           :::|||      :::   |||||||:::   ||||||   ::::::
 131 .....AspHisPhePheSerArgLeuAspArgGlnGluGlyAlaIleLeu 145

 370 ATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCA 419
     :::     |||   :::::::::::::|||          ||||||:::::
 146 PheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleThrLy 162

 420 GCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAG 469
     ::::   |||            |||:::   ||||||   |||   :::
 162 sArgTyrProGly......LeuAlaPheAlaLysProValLysAsnArgA 177

 470 CGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACC 519
       :::::::::   :::   :::   |||          :::||||||
 177 rgLeuAsnGlnLysIleIleSerLeuArgGluSerPheGlnGlyLysIle 193

 520 GCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGC 569
     :::|||         ::::::::::::|||   ::::::||||||:::
 194 ValProProGln...AsnAlaIleAsnGlnAlaLeuArgAlaLeuHisAr 209

 570 GGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAG 619
      ||||||::::::   |||:::   |||:::|||         :::|||
 209 gGlyGluValValGlyIleValGlyAspGlnValLeuLeuSerSerGluT 226

 620 GCGGCGGCGTGTGGGCGGATTTTTCGGCAAACCTGCATACACCATGACA 669
       :::               :::|||||||:::   |||:::|||   |||
 226 yrSer.........TyrProLeuPheGlySerGlnAlaPheThrThrThr 239

 670 CTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTG 719
           |||   ||||||:::                 ::::::      ::
 240 SerProAlaLeuLeuAlaTyrLysThrLysLysThrValIleAlaValAl 256

 720 CTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCG 769
     :      |||   |||:::|||:::   :::|||:::
 256 aIleTyrArgLysProAsnGlyAsnTyrLeuValVal.............. 268

 770 TCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAAC 819
                                        :::   :::|||:::
 269 ................................ProSerLysAlaPheHis 274

 820 CGCAATACCGAATATTGGATACGC........................... 843
        |||||||||      ||||||
 275 AlaAsnThrGluLeuSerIleArgGluSerThrGluGlnLeuMetAspAr 291

 844 .......CGTTTT.......................CCGACGCAGTATC 862
           ||||||                     |||   |||:::|
 291 gLeuMetArgPheLeuGluLysGlyIleThrCysLysProGluGlnTrpL 308

 863 TGTTTATGTACAACCGCTATAAA 885
     ||:::::::::::::|||:::|||
 308 euTrpLeuHisLysArgTrpLys 315
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60656

seq_documentation_block:
ID   AAB60656 standard; Protein; 291 AA.
XX
AC   AAB60656;
XX
DT   04-MAY-2001  (first entry)
XX
DE   N. meningitidis (serogroup B) MsbB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; MsbB.
XX
OS   Neisseria meningitidis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91455.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 99; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Neisseria
CC   meningitidis MsbB protein.
XX
SQ   Sequence   291 AA;


alignment_scores:
            Quality:  168.00              Length:    289
              Ratio:  1.018                 Gaps:     12
 Percent Similarity:  57.093    Percent Identity: 26.298

alignment_block:
 US-09-303-518D-571 x AAB60656   ..

 Align seg 1/1  to: AAB60656  from: 1  to: 291
```

```
 67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
    :::|||      |||::::::|||   ::::::   ||||||   |||::
  7 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysLeuAl 23

117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
    ::::      |||   ||||||:::   |||:::|||     |||
 23 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgI 40

167 GCATCGTCGCCAATATGCGGCAGGCG.............GGTTTGAAC 201
              |||:::   ::::::             |||   :::
 40 leGlyGluIleAsnLeuAlaLysCysPheProGluTrpAspGlyLysLys 56

202 CCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGG 251
      :::|||   :::::::|||      |||   :::   ||||||
 57 ArgGluThr...ValLeuLysGlnHisPheLysHisMetAlaLysLeuMe 72

252 TTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAA 301
     ||||||   :::      ::::::                ::::::::::
 72 tLeuGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerL 89

302 TGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAG 351
    ::   :::      ::::::   :::::::::::::::||||||
 89 euValArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAla 104

352 GGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
    ||||||   ::::::::::::   ||||||:::   ::::::::::::::
105 GlyGluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAl 121

402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACA 451
    :         ::::::|||::::::|||      |||   :::|||||||:
121 aValTyrAlaLeuAsnGlnAspValPro......LeuIleSerMetTyrS 136

452 AGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTG 501
    ::      |||   |||   :::|||      :::      ||||||
136 erHisGlnLysAsnLysIleLeuAspAlaGlnIleLeuLysGlyArgAsn 152

502 CGCGGCAAA..............GGCAAAACCGCGCCCACCGGCATACA 536
    |||                   |||:::|||             ::
153 ArgTyrAspAsnValPheLeuIleGlyArgThr..............Gl 164

537 AGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCA 586
    :|||||||:::   ::::::|||   :::|||   :::::|||   ::::
164 uGlyValArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheL 181

587 TC...CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC..... 624
    ::   |||||||||              |||:::   |||
181 euTyrLeuProAsp..............GlnAspPheGlyArgAsnAsp 192

625 GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGC 674
    :::|||::::::||||||||||||      :::   |||:::|||
193 SerValPheValAspPhePheGlyIleGlnThrAlaThrIleThrGlyLe 209

675 GGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCG 724
     ::::::::::|||   :::   ::::::|||::::::
209 uSerArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProV 226

725 AACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAA 774
      |||      |||:::          :::|||||||:::   |||:::
226 alArgGluAlaAspAsnThr...ValThrLeuHisPheTyrProAlaTrp 241

775 GGGGAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCG 821
      ::::::   ::::::   |||:::   ||||||      :::||||||
242 GluSerPheProSerGluAspAlaGlnAlaAspAlaGlnArgMetAsnAr 258

822 CAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGT 871
    |      |||           ::::::|||::: ||||||      |||
258 gPheIleGluGluProCysAlaAsnIleProSer.SerIlePheGlyCys 274

872 ACAACCGCTATAAAA 886
    |||::::::   |||
275 ThrSerValSerLys 279
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB88540

seq_documentation_block:
ID   AAB88540 standard; Protein; 318 AA.
XX
AC   AAB88540;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Haemophilus influenzae essential bacterial protein SEQ ID NO:98.
XX
KW   Haemophilus influenzae; essential bacterial gene; identification;
KW   otitis media; meningitis; upper respiratory tract infection;
KW   infection; antimicrobial.
XX
OS   Haemophilus influenzae.
XX
PN   WO200111033-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US21176.
XX
PR   04-AUG-1999;   99US-0368382.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Chovan LE,  Hessler PE,  Reich KA;
XX
DR   WPI; 2001-147511/15.
DR   N-PSDB; AAF94393.
XX
PT   Essential bacterial genes from Haemophilus influenzae and methods for
PT   identifying 'essential' genes that may be potential therapeutic targets
PT   -
XX
PS   Claim 9; Page 149; 185pp; English.
XX
CC   AAF94345 to AAF94409 represent essential bacterial genes from
CC   Haemophilus influenzae, which encode the proteins given in AAB88492 to
CC   AAB88556. The present invention also describes methods for identifying
CC   essential bacterial genes (i.e. those essential to the survival of a
CC   bacterium) using a transposition system. The methods are used to
CC   identify essential genes from bacteria, especially H. influenzae (which
CC   causes otitis media, meningitis and upper respiratory tract infections)
CC   which may be used as targets for potential antimicrobial agents.
CC   AAF94410 to AAF94416 represent PCR primers used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   318 AA;


alignment_scores:
            Quality:  168.00              Length:    265
              Ratio:  1.105                 Gaps:     10
 Percent Similarity:  57.358   Percent Identity: 24.151

alignment_block:
 US-09-303-518D-571 x AAB88540   ..

 Align seg 1/1  to: AAB88540  from: 1  to: 318

    112 CTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCG 161
        ||||||      :::||||||   |||           |||:::::::||
     57 LeuGlyIleTrpIleGlyHisLysAla...........LysLysGlnAr 69

    162 CGCGCGCATCGTCGCCAATATGCGGCAGGCG...........GGTTTGA 199
        |:::|||      :::|||::::::   :::             :
     69 gThrArgAlaGlnThrAsnLeuGlnTyrCysPheProHisTrpThrGluG 86

    200 ACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGC 249
        ::         |||::::::      :::||||||   :::|||:::
     86 lnGlnArgGluGlnValIleAspLysMetPheAlaValValAlaGlnVal 102
```

```
    250 GGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAAC 299
        :::   ::::::          :::::::::   :::   ::::::
    103 MetPheGlyIleGlyGluIleAlaIleArgSerLysLysHisLeuGlnLy 119

    300 AATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACA 349
              :::   :::   |||   ||||||:::::||||||      :
    119 sArgSerGluPheIle...GlyLeuGluHisIleGluGlnAlaLysAlaG 135

    350 AGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTG 399
        ::|||:::::::::::::::::::::||||||      :::   |||
    135 luGlyLysAsnIleIleLeuMetValProHisGlyTrpAlaIleAspAla 151

    400 GGCGGACGCTACATC...AGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446
        :::|||      :::   :::|||   :::|||      :::|||:::||
    152 SerGlyIleIleLeuHisThrGlnGlyMetPro......MetThrSerMe 166

    447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA 496
        ||||:::|||   :::         :::|||   :::           |
    166 tTyrAsnProHisArgAsnProLeuValAspTrpLeuTrpThrIleThrA 183

    497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA 546
        ||   |||      ||||||                :::|||:::|||
    183 rgGlnArgPheGlyGlyLysMetHisAlaArgGln...AsnGlyIleLys 198

    547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGA 596
        :::::::::   :::|||   ||||||           ||||||||
    199 ProPheLeuSerHisValArgLysGlyGluMetGlyTyrTyrLeuProAs 215

    597 CCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCG 646
        |:::      :::   ::::::      :::|||:::::|||||||||
    215 pGluAspPheGlyAlaGluGln......SerValPheValAspPhePheG 230

    647 GCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC... 693
        ||            |||:::           |||:::|||   :::
    230 lyThrTyrLysAlaThrLeuProGlyLeuAsnLysMetAlaLysLeuSer 246

    694 AAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACA 743
        |||:::|||      :::|||      :::        |||::
    247 LysAlaValValIleProMetPheProArgTyrAsnAlaGluThrGlyLy 263

    744 AGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACA 793
        :   :::   ::::::|||:::|||::::::   :::|||:::   :::
    263 s...TyrGluMetGluIleHisProAlaMet...AsnLeuSerAspAspP 278

    794 AAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGC 843
            ::::::|||   :::::::|||::::::   |||   ::::::
    278 roGluGlnSerAlaArgAlaMetAsnGluGluIleGluSerPheValThr 294

    844 CGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACG 888
            |||   ||||||:::::::::   :::      :::|||
    295 ProAlaProGluGlnTyrValTrpIleLeuGlnLeuLeuArgThr 309

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60654

seq_documentation_block:
ID   AAB60654 standard; Protein; 318 AA.
XX
AC   AAB60654;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Haemophilus influenzae MsbB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; MsbB.
XX
OS   Haemophilus influenzae.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91453.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 98-99; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Haemophilus
CC   influenzae MsbB protein.
XX
SQ   Sequence   318 AA;


alignment_scores:
           Quality:  168.00              Length:    265
             Ratio:  1.105                 Gaps:     10
 Percent Similarity:  57.358   Percent Identity: 24.151

alignment_block:
 US-09-303-518D-571 x AAB60654   ..

 Align seg 1/1  to: AAB60654  from: 1  to: 318

    112 CTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCG 161
        ||||||      :::||||||   |||            |||::::::||
     57 LeuGlyIleTrpIleGlyHisLysAla.....LysLysGlnAr 69

    162 CGCGCGCATCGTCGCCAATATGCGGCAGGCG.....GGTTTGA 199
        |:::|||      :::|||::::::   :::           :
     69 gThrArgAlaGlnThrAsnLeuGlnTyrCysPheProHisTrpThrGluG 86

    200 ACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGC 249
        ::         |||::::::      :::|||||||   :::|||:::
     86 lnGlnArgGluGlnValIleAspLysMetPheAlaValValAlaGlnVal 102
```

```
 250 GGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAAC 299
     :::   ::::::         :::::::::   :::   ::::::
 103 MetPheGlyIleGlyGluIleAlaIleArgSerLysLysHisLeuGlnLy 119

 300 AATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACA 349
           :::   :::   |||   ||||||::::::||||||      :
 119 sArgSerGluPheIle...GlyLeuGluHisIleGluGlnAlaLysAlaG 135

 350 AGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTG 399
     ::|||:::::::::::::::::::::||||||      :::   ||||
 135 luGlyLysAsnIleIleLeuMetValProHisGlyTrpAlaIleAspAla 151

 400 GGCGGACGCTACATC...AGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446
     :::|||      :::   :::|||   :::|||      :::|||:::||
 152 SerGlyIleIleLeuHisThrGlnGlyMetPro......MetThrSerMe 166

 447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA 496
     ||||:::|||   :::         :::|||   :::            |
 166 tTyrAsnProHisArgAsnProLeuValAspTrpLeuTrpThrIleThrA 183

 497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA 546
     ||   |||      ||||||                  :::|||:::|||
 183 rgGlnArgPheGlyGlyLysMetHisAlaArgGln...AsnGlyIleLys 198

 547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGA 596
     :::::::::   :::|||   ||||||           ||||||||
 199 ProPheLeuSerHisValArgLysGlyGluMetGlyTyrTyrLeuProAs 215

 597 CCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCG 646
     |:::      :::   ::::::      :::|||::::::|||||||||
 215 pGluAspPheGlyAlaGluGln......SerValPheValAspPhePheG 230

 647 GCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC... 693
     ||           |||:::            |||:::|||   :::
 230 lyThrTyrLysAlaThrLeuProGlyLeuAsnLysMetAlaLysLeuSer 246

 694 AAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACA 743
     |||:::|||         :::|||          :::         |||::
 247 LysAlaValValIleProMetPheProArgTyrAsnAlaGluThrGlyLy 263

 744 AGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACA 793
     :   :::   :::::::|||:::|||:::::   :::|||:::   :::
 263 s...TyrGluMetGluIleHisProAlaMet...AsnLeuSerAspAspP 278

 794 AAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGC 843
           :::::::|||   ::::::|||::::::   |||   ::::::
 278 roGluGlnSerAlaArgAlaMetAsnGluGluIleGluSerPheValThr 294

 844 CGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACG 888
           |||   ||||||::::::::::   :::       :::|||
 295 ProAlaProGluGlnTyrValTrpIleLeuGlnLeuLeuArgThr 309

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60655

seq_documentation_block:
ID   AAB60655 standard; Protein; 347 AA.
XX
AC   AAB60655;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Moraxella catarrhalis MsbB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; MsbB.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91454.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 99; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Moraxella catarrhalis
CC   MsbB protein.
XX
SQ   Sequence   347 AA;

alignment_scores:
            Quality:  151.00              Length:    305
              Ratio:  0.944                Gaps:     13
 Percent Similarity:  52.459   Percent Identity: 25.246

alignment_block:
 US-09-303-518D-571 x AAB60655   ..

 Align seg 1/1  to: AAB60655  from: 1  to: 347

  55 CTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCT 104
     |||      ||||||:::      |||   :::|||   |||
  52 LeuAlaPheAlaLeuIleLeuProLeuIlePheLeuProLeuArgTrpGl 68

 105 GCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGG 154
           :::|||:::||||||||||   |||:::   ||||||   |||:
  68 nPheTrpIleGlyLysArgLeuGlyIleLeuValHisTyrLeuAlaLysS 85

 155 AAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAAC... 201
     ::              ::::::|||::::::   :::      |||
  85 erArgValGlnAspThrLeuThrAsnLeuGlnLeuThrPheProAsnGln 101
```

```
 202 CCCGACACGCAG.........ACGGTCAAAGCCGTTTTTGCGGAAACGGC 242
 102 ProLysSerLysHisLysAlaThrAlaArgGlnValPheIleAsnGlnGl 118

 243 AAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACA 292
 118 yIleGlyIlePheGluSerLeuCysAlaTrpPhe...ArgProAsnValP 134

 293 TCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCT 342
 134 heLysArgThrPhe...SerIleSerGlyLeuGlnHisLeuIleAspAla 149

 343 TTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTA 392
 150 GlnLysGlnAsnLysAlaValIleLeuLeuGlyGlyHisArgThrThrLe 166

 393 CGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCG 442
 166 uAspLeuGlyGlyArgLeuCysThrGlnPhe......PheAlaAlaAspC 181

 443 CCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCG 492
 181 ysValTyrArgProGlnAsnAsnProLeuLeuGluTrpPheIleTyrAsn 197

 493 GGCAGG.....................GTGCGCGGCAAAGGCAAAACCGC 521
 198 AlaArgArgCysIlePheAspGluGlnIleSerAsnArg........... 210

 522 GCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGG 571
 211 ...............AspMetLysLysLeuIleThrArgLeuLysGlnG 222

 572 GCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGC 621
 222 lyArgIleIleTrpTyrSerProAsp..............GlnAspPhe 233

 622 GGC.........GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACAC 662
 234 GlyLeuGluHisGlyValMetAlaThrPhePheGlyValProAlaAlaTh 250

 663 CATGACACTGGCGGCAAAATTGGCACACGTCAA....AGGCGTGAAAACC 708
 250 rIleThrAlaGlnArgArgLeuIleLysLeuGlyAspLysAlaAsnProP 267

 709 CTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCA 758
 267 roValIleIleMetMetAspMetLeuArgGlnThrProAspTyrIleAla 283

 759 ......CATCCGCCCC.........................GTCCAAG 775
 284 LysGlyHis.ArgProHisTyrHisIleSerLeuSerAlaValLeuLysA 300

 776 GGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAAT 825
 300 snTyrProSerAspAspGluThrAlaAspAlaGluArgIleAsnArgLeu 316

 826 ACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAA 875
 317 IleGluGlnAsnIleGlnLysAspLeuThrGlnTrpMetTrpPheHisAr 333

 876 CCGCTATAAAACG 888
 333 gArgPheLysThr 337
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB65602

seq_documentation_block:
ID ABB65602 standard; Protein; 954 AA.
XX
AC ABB65602;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 23598.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL09705.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 23598; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 954 AA;

alignment_scores:
Quality: 145.00 Length: 329
Ratio: 1.074 Gaps: 17
Percent Similarity: 41.033 Percent Identity: 25.228

alignment_block:
US-09-303-518D-571 x ABB65602 ..

Align seg 1/1 to: ABB65602 from: 1 to: 954

```
 150 AAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGG..... 194
 347 ArgGlyAsnAspArgGly......ArgGlnPheGlyGlyGlyGlyGlyGl 361

 195 ..................TTTGAACCCCGACACGCAGAC......GGTC 219
 361 yArgArgGlyGlyArgArgPheSerProArgGlyArgSerProMetArgG 378

 220 AAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTT............... 254
 378 lnAspAsnGlyGlyGlyAsnGlyGlyAsnArgPheGlnArgSerAsnSer 394

 255 ..........................GGAACTTGCCCCCGCGTTTTTCA 277
 395 ArgArgArgSerArgSerArgArgLeuSerArgSerProMetArgTyrSe 411

 278 AAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAA 327
```

```
    411 rArgSerProArgArgPheAsnAsnArgArgArgSerProMetMetAsnP 428

    328 CACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCC 377
          |||::::::|||   |||   ||||||                 :::
    428 heArgGlyGlyGlyArgGlyGlyGlyArg.............GlyGly 439

    378 GCACATCGGCAGCTA..........................CGATT 397
        :::||||||||||:::                          |||
    440 GlyHisArgGlnMetTrpGlnHisArgGlyGlySerProAsnPheArgGl 456

    398 TGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCT..... 437
         |||      :::|||      |||         |||
    456 yGlyGlyArgIleHisTrpGlnAlaArgHisSerProGlyGlyGlyGlyA 473

    438 ..........................GACCGCCATGTACAAGCCGCC 458
                                   ||||||
    473 rgGlyGlyGlyGlyGlyMetGlyArgPheAspArg............. 484

    459 GAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCA 508
          :::||||||   :::||||||
    485 ...ArgGlnSerProGlnGlnAsn....................... 491

    509 AAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAAT..... 551
         |||   ||||||   ||||||::::::      |||   |||
    492 .ArgTyrAsnArgAspHisArgGlnSerProMetGlnGlnAsnGlnProP 508

    552 ...................CATCAAGGCCCTGCGCGCGGGCGAGGCAAC 581
                            ||||||||||||:::|||      ::::::
    508 heArgLysGlnPheSerProHisGlnGlyProGlyArgArgPheSerSer 524

    582 CATCATCCTGCCCGA............CCACGTCCCTTCTCCGCAGGAAG 619
          :::     |||          ||||||   |||      ||
    525 ProGlnGlnArgArgAsnSerArgAspArgArgProAsnSerArgGluAr 541

    620 GCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACA 669
        ||||         ||||||
    541 gArgSerSerProGlyGly............................. 547

    670 CTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTG 719
          ||||||      |||      :::|||  :::||||||
    548 ..GlyGly......GlyHisMetAsnArgTrpAspAsnPro......... 558

    720 CTGCGAACGCCTGCC............................... 734
                ||||||
    559 ......ProProAlaArgAsnArgArgSerSerSerGlySerSerAlaGl 573

    735 ....CGACGGACAAGGCTTCGTGTTGCACATCCGCCC............. 767
            |||:::   |||   |||   |||   ||||||
    573 yGlyArgGlnGlnArgGlnArgSerAlaSerProProGluLysArgArgS 590

    768 ..........CGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGC 806
                 |||   |||:::::::::|||::::::|||   |||
    590 erSerHisSerArgSerArgSerLeuSerArgAsnArgSer...ArgSer 605

    807 CGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGC 856
        |||      :::   :::   |||      ::::::|||::::::   ::
    606 ArgSerSerSerGlnHisSerArgLysArgGluSerProIleGlyArgSe 622

    857 AGTATCTGTTTATGTACAACCGCTATAAAACGCCGTA 893
        :       |||:::   |||      |||::::::
    622 rSerValGluTyrAlaGlyProAlaValAsnThrIle 634

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813

seq_documentation_block:
ID   AAB59813 standard; Protein; 1017 AA.
XX
AC   AAB59813;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #4.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1017 AA;


alignment_scores:
            Quality:  145.00              Length:    358
              Ratio:  1.090                Gaps:     17
 Percent Similarity:  37.151   Percent Identity: 24.022

alignment_block:
 US-09-303-518D-571 x AAB59813   ..

 Align seg 1/1  to: AAB59813  from: 1  to: 1017

      41 CCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTG 90
         |||   |||:::   |||   |||   |||      :::   ||||||
      11 ProThrCysSerArgCysIleProAsnCysPro...ThrTrpProCys.. 25

      91 TCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTT 140

      25 ................................................. 25

     141 TTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGG 190
                    |||   |||                       ||||||
      26 ...........ArgThrThr.....................CysGly... 30

     191 CGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACG 240
                       |||   |||||||||:::|||      |||:::||
      31 ............AlaThrThrArgArgSerArgProThrArgArgArgAr 43

     241 G...CAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAAC.... 283
         |   :::|||:::      ::::::      |||   |||   :::
      43 gSerMetAsnThrGlySerArgIleAlaCysArgAlaSerValSerProI 60

     284 ........CGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGG 325
                 |||:::||||||      |||   :::
      60 leSerIleArgGlnThrSerAlaAlaCysIleArg.............. 71
```

```
 326 AACACGTGCAGCAGGCTTTGGACAAGG.....................GC 354
                ||| |||  |||                          ::
  72 .........SerAlaAlaTrpArgArgArgProSerLeuProValThrTh 85

 355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG.... 400
     :  :::|||      ::::::  :::      :::      |||      .
  85 rAlaSerCysArgProThrLysProSerTrpLysThrGlyCysTrpArgA 102

 400 ................................................. 400

 102 laSerSerSerProLysSerIleSerProLysProArgProThrCysArg 118

 401 ..............GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCAC 435
                   |||      :::|||::::::|||   ||||||||
 119 ProSerProGlyThrAlaArgArgValSerThrThrSerProArgSerTh 135

 436 CTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCAT 485
     |            |||:::|||||||   |||:::         :::|||:
 135 r............ThrGlyArgArgTrpSerSerProAlaArgArgSerA 148

 486 GCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATAC 535
     ::   |||||||||   |||:::   |||
 148 laGlyArgAlaGlyArgAlaGlyCysAla..................... 157

 536 AAGGGGTCAAACAAATCATCAAGGCCC..................... 562
     ::::::|||:::|||:::|||||||||
 158 ArgSerSerArgLysThrSerArgProIleArgSerAlaArgProSerCy 174

 562 ................................................. 562

 174 sSerLysSerProThrSerValSerAlaPheProProSerProAlaArgA 191

 563 ..............TGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACC 598
                  |||   |||         |||||||||
 191 laSerArgThrArgCysArgArgAsnSerLeuProSerSerValThrArg 207

 599 ACGTCCCTTCTC..............CGCAGGAAGGCGGCGGCGTGTGG 633
     :::|||                   |||||||||:::   :::|||||
 208 SerSerAlaThrArgAlaAlaThrProArgArgLysThrProCysCysGl 224

 634 GCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATT 683
     ||||                 |||          |||   :::|
 224 yArgThrThrArgProProSerSerThrArgAsnSerSerArgAlaThrT 241

 684 GGCACACGTCAAAGGCGTGAAAA..............CCCTGTTTTTCT 718
     ||         ::::::   :::              |||   ::::::
 241 rpMetArgTrpAsnSerSerArgTrpAsnValArgPheProSerMetAla 257

 719 GCTGCGAACGCCTGCCCGACGGACAAGGCTTCG............TGTTG 756
        |||:::   :::||||||   |||:::|||              ||
 258 ProAlaSerArgAlaProThrAlaLysSerSerArgGlyArgThrIleCy 274

 757 CACATCCGCCCCGTCC..............AAGGGGAATTGAACGGCAA 791
     |:::|||:::||||||             :::::::::      ||||
 274 sSerSerSerProSerAlaAlaProThrProArgAlaArgThrProAlaT 291

 792 CAAAGCCCACG.................................ATG 805
     ||   ||||||
 291 hrThrProThrProSerSerArgGlnProSerGlySerAlaArgProSer 307

 806 CCGCCGTGTTCAACCGCAATACCG 829
     ||||||   |||:::|||||||||
 308 ProProSerSerSerAlaIlePro 315
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826

seq_documentation_block:

```
ID   AAB59826 standard; Protein; 1615 AA.
XX
AC   AAB59826;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #3 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872,
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1615 AA;
```

alignment_scores:

```
            Quality:  145.00              Length:    358
              Ratio:  1.090                 Gaps:     17
 Percent Similarity:  37.151   Percent Identity: 24.022
```

alignment_block:

US-09-303-518D-571 x AAB59826 ..

Align seg 1/1 to: AAB59826 from: 1 to: 1615

```
  41 CCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTG 90
     |||   |||:::  |||   |||   |||      :::   ||||||
 609 ProThrCysSerArgCysIleProAsnCysPro...ThrTrpProCys.. 623

  91 TCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTT 140

 623 ................................................. 623

 141 TTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGG 190
                |||   |||                      ||||||
 624 ...........ArgThrThr...................CysGly... 628

 191 CGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACG 240
                     |||   |||||||||:::|||      |||:::||
 629 ...........AlaThrThrArgArgSerArgProThrArgArgArgAr 641

 241 G...CAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAAC.... 283
     |   :::|||:::      ::::::      |||   |||   :::
 641 gSerMetAsnThrGlySerArgIleAlaCysArgAlaSerValSerProI 658
```

```
 284 ........CGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGG 325
             |||:::||||||      |||   :::
 658 leSerIleArgGlnThrSerAlaAlaCysIleArg.............. 669

 326 AACACGTGCAGCAGGCTTTGGACAAGG.....................GC 354
             |||      |||   |||                      ::
 670 .........SerAlaAlaTrpArgArgArgProSerLeuProValThrTh 683

 355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG.... 400
     :   :::|||      ::::::   :::      :::      |||
 683 rAlaSerCysArgProThrLysProSerTrpLysThrGlyCysTrpArgA 700

 400 ..................................................... 400

 700 laSerSerSerProLysSerIleSerProLysProArgProThrCysArg 716

 401 ..............GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCAC 435
                   |||      :::|||:::::::|||   ||||||||
 717 ProSerProGlyThrAlaArgArgValSerThrThrSerProArgSerTh 733

 436 CTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCAT 485
     |            |||:::||||||   |||:::         :::|||:
 733 r............ThrGlyArgArgTrpSerSerProAlaArgArgSerA 746

 486 GCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATAC 535
     ::   |||||||||   |||:::   |||
 746 laGlyArgAlaGlyArgAlaGlyCysAla..................... 755

 536 AAGGGGTCAAACAAATCATCAAGGCCC..................... 562
     ::::::|||:::|||:::|||||||||
 756 ArgSerSerArgLysThrSerArgProIleArgSerAlaArgProSerCy 772

 562 ..................................................... 562

 772 sSerLysSerProThrSerValSerAlaPheProProSerProAlaArgA 789

 563 ..............TGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACC 598
                   |||   |||         |||||||||
 789 laSerArgThrArgCysArgArgAsnSerLeuProSerSerValThrArg 805

 599 ACGTCCCTTCTC..............CGCAGGAAGGCGGCGGCGTGTGG 633
     :::|||                   |||||||||:::   :::|||||
 806 SerSerAlaThrArgAlaAlaThrProArgArgLysThrProCysCysGl 822

 634 GCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATT 683
     ||||                  |||            |||   :::|
 822 yArgThrThrArgProProSerSerThrArgAsnSerSerArgAlaThrT 839

 684 GGCACACGTCAAAGGCGTGAAAA..............CCCTGTTTTTCT 718
     ||         ::::::   :::               |||   ::::::
 839 rpMetArgTrpAsnSerSerArgTrpAsnValArgPheProSerMetAla 855

 719 GCTGCGAACGCCTGCCCGACGGACAAGGCTTCG............TGTTG 756
        |||:::   :::||||||   |||:::|||             ||
 856 ProAlaSerArgAlaProThrAlaLysSerSerArgGlyArgThrIleCy 872

 757 CACATCCGCCCCGTCC..............AAGGGGAATTGAACGGCAA 791
     |:::|||:::||||||             :::::::::      ||||
 872 sSerSerSerProSerAlaAlaProThrProArgAlaArgThrProAlaT 889

 792 CAAAGCCCACG.................................ATG 805
     ||   ||||||
 889 hrThrProThrProSerSerArgGlnProSerGlySerAlaArgProSer 905

 806 CCGCCGTGTTCAACCGCAATACCG 829
     ||||||   |||:::|||||||||
 906 ProProSerSerSerAlaIlePro 913

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955

seq_documentation_block:
ID   AAY04955 standard; Protein; 573 AA.
XX
AC   AAY04955;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 41T#3.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34206.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 41T; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   573 AA;

alignment_scores:
            Quality:  141.00            Length:    350
              Ratio:  0.966               Gaps:     17
 Percent Similarity:  41.714   Percent Identity: 23.714

alignment_block:
 US-09-303-518D-571 x AAY04955   ..

 Align seg 1/1  to: AAY04955  from: 1  to: 573

      36 GCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCC 85
         |||:::||||||::::::|||:::   |||||||||
     254 AlaAspArgHisGlyTyrProThrProArgProAlaIleArgGlyAspVa 270

      86 TGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACCG......GCTCGGA 129
           :::|||:::   :::::::::   :::|||   |||      :::   |
     270 lSerValGlyGlyMet***CysCysSerGlyGlyProValAlaGlySerT 287

     130 CATCTGGCGTTTTACCTTTTAAAGGA..................... 155
         ||:::|||:::           |||
     287 hrGlnGlyIleGly***ValGlyGlyHisArgArgCysSerAlaArgGln 303

     156 ......AGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGA 199
               |||   |||   |||||||||:::   :::      |||   ::
     304 LeuLeuArgThrArgProHisArgArgArgArgCysArgArgGlySerAr 320

     200 A....................................................... 200
         :
     320 gIleGlyGlyGlyAlaSer***ProAspArgAspLeuGlyAlaArgPheA 337
```

```
201 .....................................CCCCGACACGCA 212
                                        ||||||:::|||
337 rgAspGlnArgIleAlaGlyArgTrpLeuAspAlaGlyProArgArgAla 353

213 GACGGTCAAAGCCGT.................TTTTGCGGAAACGGCAA 244
      |||:::   |||                        :::|||::
354 GlyGlyArgArgArgArgArgCysArgArgAlaValArgArgGlyGlyAr 370

245 AATGCGGTTTGGAACT.............................. 260
    ::::|||   :::|||
370 gLeuArgAlaAlaThrGlySerArgArgArgAspThrGlyArgArgTyrG 387

261 ..TGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAA 308
      ||||||   :::            ||||||||||||:::::::::::
387 lnCysProProAlaGlyAlaGlyArgGlyArgHisArgArgArgAlaArg 403

309 AGCGGTACA....................................... 317
    :::|||:::
404 AspGlyAlaAlaGlnTrpLeuCysGlyArgArgArgThrGlyGlyArgVa 420

318 .......................CGGCTGGGAACACGTGCA....... 335
                            |||   ||||||||||||
420 lTyrArgGlyAspArgLeuGlyArgArgArgGlyThrArgAlaAspArgI 437

336 ........GCAGGCTTTGGACAA........................... 350
           ||||||   |||:::
437 leAspGlyAlaGlyValGlyArgAlaGlyArgAla***ArgGlyProPro 453

351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAG.........CT 391
    |||||||||      ::::::|||:::      |||:::
454 GlyArgArgArgArgLeuGlnHisGlyProCysArgArgCysPheProAl 470

392 ACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACC 441
     |||:::|||         |||      :::|||::::::         |
470 aArgIleGlyAlaHisCysHisProLysGlyAlaAspLeu......GlyA 485

442 GCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGC 491
    ||:::::::||||||:::::             |||:::::::::   |||
485 rgTyrLeuGlnAlaGlyArg............ArgSerGlyTyrArgGly 497

492 GGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGG 541
      :::||||||   :::|||                |||:::   |||
498 ArgArgGlyAlaAspArgArg..............ArgArgCysArg.. 508

542 TCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTG 591
                                  ||||||:::|||:::   :
509 ............................ArgGlyGlyHisArgSerG 515

592 CCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTT 641
    ::||||||         :::   |||||||||         |||:::
515 lyArgPro...ValValGlyIleGlyArgArgSerGlyAspGlyAlaAsn 530

642 TTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACG 691
    :::|||:::        ::::::     |||   :::   ||||||
531 TrpArgArgArg......AsnArgArgArgGlyCysArgProGlyThrAl 545

692 TCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGA...C 738
                                |||   |||:::|||   |
545 a......................CysAlaArgProProSerArgHisA 554

739 GGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGG 788
    ||:::   |||         :::   ||||||   |||:::   :::|||
554 rgAlaGlyLeuLeuProHisArgThrProArgArgArgAlaAlaAspArg 570

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60653

seq_documentation_block:
ID   AAB60653 standard; Protein; 311 AA.
XX
AC   AAB60653;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Haemophilus influenzae HtrB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; HtrB.
XX
OS   Haemophilus influenzae.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91452.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 98; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Haemophilus
CC   influenzae HtrB protein.
XX
SQ   Sequence   311 AA;


alignment_scores:
           Quality:  140.00              Length:     291
             Ratio:  0.897                Gaps:       9
 Percent Similarity:  53.608   Percent Identity: 23.368

alignment_block:
 US-09-303-518D-571 x AAB60653   ..

 Align seg 1/1  to: AAB60653  from: 1  to: 311
```

```
 64 GCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT 113
    |||:::   :::   :::      |||           |||::: . ::
 26 AlaIleTrpArgSerIleLeuCysLeuProTyrProIleLeuArgHisIl 42

114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
    :|||:::   :::|||   |||      :::|||            |||
 42 eGlyHisGlyPheGlyTrpLeuPheSerHisLeuLysValGlyLysArgA 59

164 CGCGCATCGTCGCC.............................. 177
          :::|||
 59 rgAlaAlaIleAlaArgArgAsnLeuGluLeuCysPheProAspMetPro 75

178 ...AATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGC 224
       |||   |||::::::   |||:::   :::   :::::::|||
 76 GluAsnGluArgGluThrIleLeuGlnGluAsnLeuArgSerValGlyMe 92

225 CGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTT 274
     ::::::   ||||||:::                        |||:::|
 92 tAlaIleIleGluThrGlyMet...................AlaTrpP 102

275 TCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGG 324
    ||   :::   :::   |||:::         |||   |||:::|||
102 heTrpSerAspSerArgIleLysLysTrpSerLys...ValGluGlyLeu 117

325 GAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCAC 374
    ::::::::::::::      ::::::   :::|||:::::::::::::
118 HisTyrLeuLysGlu......AsnGlnLysAspGlyIleValLeuValGl 132

375 GCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGC 424
       |||:::   :::   :::|||||:::|||   |||:::   :::
132 yValHisPheLeuThrLeuGluLeuGlyAlaArgIleIleGlyLeuHisH 149

425 TTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATA 474
      |||            :::::::|||:::|||                :::
149 isProGly......IleGlyValTyrArgProAsnAspAsnProLeuLeu 163

475 GACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCC 524
    |||   ::::::      ||||||:::|||::::::
164 AspTrpLeuGlnThrGlnGlyArgLeuArgSerAsnLysAspMetLeuAs 180

525 CACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCG 574
              :::   ::::::   :::|||||||||||||||      |
180 pArg......LysAspLeuArgGlyMetIleLysAlaLeuArgHisGluG 195

575 AGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC 624
    ||:::            |||||||||      :::      :::   :::
195 luThrIleTrpTyrAlaProAspHis......AspTyrGlyArgLysAsn 209

625 GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGC 674
    :::|||::::::   ||||||:::   |||           |||   ::
210 AlaValPheValProPhePheAlaValProAspThrCysThrThrThrGl 226

675 GGCA......AAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCT 718
    ::::         |||         ::::::   |||::::::   |||:
226 ySerTyrTyrLeuLeuLysSerSerGlnAsnSerLysValIleProPheA 243

719 GCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATC...CGC 765
    ::      |||      ||||||:::|||:::::::::   |||
243 laProLeuArgAsnLysAspGlySerGlyTyrThrValSerIleSerAla 259

766 CCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTT 815
    ||||||:::                :::::::::      ||||||   ::
260 ProValAspPheThrAspLeuGlnAspGluThrAlaIleAlaAlaArgMe 276

816 CAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGT 865
    :|||:::   :::|||      |||   :::      :::||||||::::
276 tAsnGlnIleValGluLysGluIleMetLysGlyIleSerGlnTyrMetT 293

866 TTATGTACAACCGCTATAAAACG 888
    ::::::::::|||:::||||||
293 rpLeuHisArgArgPheLysThr 300
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34697

seq_documentation_block:
ID   AAY34697 standard; Protein; 463 AA.
XX
AC   AAY34697;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Chlamydia pneumoniae lipoprotein sequence.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 699-700; Disclosure; 1912pp; English.
XX
CC   AAY34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   463 AA;


alignment_scores:
           Quality:  137.00              Length:    338
             Ratio:  0.867                 Gaps:     15
 Percent Similarity:  46.746   Percent Identity: 23.964

alignment_block:
 US-09-303-518D-571 x AAY34697   ..

 Align seg 1/1  to: AAY34697  from: 1  to: 463

      52 ATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTG 101
         ||||||   :::   |||         :::|||   :::         ||
      13 IleLeuGluAlaProLeuTyrTyrLeuValSerGlyIleIleAlaLeuCy 29

     102 TCTGCACACG.....................CTGGGAAACCGGCTCGGAC 130
         |   ||||||                    ||||||:::   :::|||
      29 sArgHisThrProArgSerPheLeuThrGlyLeuGlyLysGlyPheGlyP 46

     131 ATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180
            ||||||||||||:::::::::::   |||         ::::::|||
      46 heLeuAlaPheTyrIleIleSerAspTyrArgLysThrAlaLeuThrAsn 62
```

```
181 ATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTT 230
    :::      |||                          :::   ::::||
 63 LeuAlaLeuAla.....................PheProGluLysThrPh 72

231 TGCGGAAACGGCAAAATGCGGT.......................... 252
    |   |||      |||   :::
 72 eAspGluArgTyrLysIleAlaArgGlnSerLeuGlnHisLeuIleIleT 89

253 .....TTGGAACTTGCC.............................. 264
         |||||||||
 89 hrLeuLeuGluLeuLeuAlaIleGluGlnLeuValGlyAsnIleAspLys 105

265 ............................CCCGCGTTTTTCAAAAAACC 284
                                   |||      |||::::::
106 LeuIleThrIleValThrSerSerArgAsnProLysGlyPheSerSerGl 122

285 G..............GAAGACATCGAAACAATGTTCAAAGCGGTACACG 319
                  ||||||:::|||      ||||||
122 uGluValIleSerAsnGluAspLeuGluGluThrPheLys.......... 135

320 GCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTC 369
                        |||::::::   :::||||||:::::::
136 ...................AsnLeuGlnGluLysGlnGlyLeuIleLeu 145

370 ATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCA 419
    :::      |||   :::::::::::|||         ||||||:::::
146 PheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleThrLy 162

420 GCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAG 469
    ::::   |||            |||:::   |||      |||   :::
162 sAsnTyrProGly......IleAlaPheAlaLysAlaIleLysAsnGlnA 177

470 CGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACC 519
      :::::::|||   :::   |||   |||          |||||||||
177 rgLeuSerLysLysIlePheAlaLeuArgGluValPheLysGlyLysIle 193

520 GCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGC 569
    :::|||         :::|||:::::::|||   |||:::|||||||:::
194 ValProProLys...AsnGlyIleGlnGlnGlyIleGluAlaLeuAsnGl 209

570 GGGCGAGGCAACCATCATCCTGCCCGACCACGTC................ 603
     |||:::   :::   |||:::   |||::::::
209 nGlyLysLeuValGlyIleValGlyAspGlnAlaLeuLeuMetSerSerT 226

604 ........CCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTC 645
            |||                                 :::|||
226 yrThrTyrPro.................................LeuPhe 231

646 GGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAA 695
    |||:::||||||:::|||   |||      |||   |||||||:::
232 GlySerProAlaPheThrThrThrSerProAlaLeuLeuAlaTyrLysTh 248

696 AGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAG 745
     |||      ::::::            :::|||      :::::::::|
248 rGlyPheProValIleAlaValAsnValSerArg......GlnAlaLysG 263

746 GCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAA 795
    |||||      ::::::   |||   ::::::::::|||   :::||||||
263 lyPhe......GluValIlePro...SerAlaLysLeuTyrAlaAsnLys 276

796 GCC............CACGATGCCGCCGTGTTCAACCGCAAT........ 825
    :::            :::::::::|||::::::      :::
277 SerLeuProMetLysGluSerValAlaIleLeuMetAspGlnMetMetGl 293

826 ....ACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGT 871
          |||      |||         |||   |||:::::::::::::
293 yPheLeuGluLysGlyIleAlaSerGlnProGluGlnTrpMetTrpIleH 310

872 ACAACCGCTATAAA 885
    :::::|||:::|||
310 isLysArgTrpLys 314
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW25084

```
seq_documentation_block:
ID   AAW25084 standard; Protein; 311 AA.
XX
AC   AAW25084;
XX
DT   30-DEC-1997  (first entry)
XX
DE   Haemophilus influenzae htrB polypeptide.
XX
KW   Vaccine; htrB gene; Gram-negative bacterium; non-toxic mutant;
KW   pathogen; endotoxin; diagnosis; passive immunisation.
XX
OS   Haemophilus influenzae strain 2019.
XX
PN   WO9719688-A1.
XX
PD   05-JUN-1997.
XX
PF   27-NOV-1996;   96WO-US18984.
XX
PR   01-DEC-1995;   95US-0565943.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (REGC ) UNIV CALIFORNIA.
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Apicella MA,  Arumugham R,  Gibson BW,  Lee N,  Sunshine MG;
XX
DR   WPI; 1997-310355/28.
DR   N-PSDB; AAT79708.
XX
PT   New Gram-negative bacterial pathogen vaccines - comprising a htrB
PT   mutant or an endotoxin isolated from an htrB mutant optionally
PT   conjugated to a carrier protein.
XX
PS   Example 1; Page 61-62; 79pp; English.
XX
CC   This polypeptide comprises the htrB gene product (see also AAT79708)
CC   of Haemophilus influenzae strain 2019.  A claimed vaccine
CC   formulation contains as an active ingredient an htrB mutant of a
CC   Gram-negative bacterial pathogen (GNBP), endotoxin isolated from an
CC   htrB mutant (A) of a GNBP, endotoxin isolated from (A) conjugated
CC   to a carrier protein, or (A) which has been genetically engineered
CC   to express at least one heterologous vaccine antigen, where (A)
CC   lacks one or more secondary acyl chains of lipid A contained in the
CC   GNBP resulting in reduced toxicity when compared to lipid A of the
CC   GNBP.  Also claimed is a method for producing endotoxin-specific
CC   antisera for diagnostic assays, or for passive immunisation,
CC   comprising immunising an individual with a vaccine formulation
CC   comprising an active ingredient as above, and collecting antibodies
CC   produced from the immunised individual.
XX
SQ   Sequence   311 AA;


alignment_scores:
           Quality:  136.00              Length:    291
             Ratio:  0.872                 Gaps:      9
 Percent Similarity:  53.608   Percent Identity: 23.024

alignment_block:
 US-09-303-518D-571 x AAW25084   ..

 Align seg 1/1  to: AAW25084  from: 1  to: 311

     64 GCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT 113
        |||:::   :::   :::      |||           |||:::   ::
     26 AlaIleTrpArgSerIleLeuCysLeuProTyrProIleLeuArgHisIl 42

    114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
        :|||:::   :::|||   |||      :::|||            |||
     42 eGlyHisGlyPheGlyTrpLeuPheSerHisLeuLysValGlyLysArgA 59
```

```
    164 CGCGCATCGTCGCC..................................... 177
                ::::|||
     59 rgAlaAlaIleAlaArgArgAsnLeuGluLeuCysPheProAspMetPro 75

    178 ...AATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGC 224
           |||   |||::::::   |||:::   :::   ::::::|||
     76 GluAsnGluArgGluThrIleLeuGlnGluAsnLeuArgSerValGlyMe 92

    225 CGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTT 274
          ::::::   ||||||:::                          |||:::|
     92 tAlaIleIleGluThrGlyMet.....................AlaTrpP 102

    275 TCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGG 324
        ||   :::   :::   |||:::         |||   |||:::|||
    102 heTrpSerAspSerArgIleLysLysTrpSerLys...ValGluGlyLeu 117

    325 GAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCAC 374
        ::::::::::::::      ::::::   :::|||:::::::::::::
    118 HisTyrLeuLysGlu......AsnGlnLysAspGlyIleValLeuValGl 132

    375 GCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGC 424
           |||:::   :::   :::|||||||:::|||   |||:::   :::
    132 yValHisPheLeuThrLeuGluLeuGlyAlaArgIleIleGlyLeuHisH 149

    425 TTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATA 474
          |||            ::::::|||:::|||                ::::
    149 isProGly......IleGlyValTyrArgProAsnAspAsnProLeuLeu 163

    475 GACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCC 524
        |||   ::::::      ||||||:::|||::::::
    164 AspTrpLeuGlnThrGlnGlyArgLeuArgSerAsnLysAspMetLeuAs 180

    525 CACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCG 574
                ::::   ::::::   :::|||||||||||||||      |
    180 pArg......LysAspLeuArgGlyMetIleLysAlaLeuArgHisGluG 195

    575 AGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC 624
        ||:::            |||||||||      :::      :::   :::
    195 luThrIleTrpTyrAlaProAspHis......AspTyrGlyArgLysAsn 209

    625 GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGC 674
        :::|||::::::   ||||||:::   |||            |||   ::
    210 AlaValPheValProPhePheAlaValProAspThrCysThrThrThrGl 226

    675 GGCA......AAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCT 718
        ::::         |||         ::::::   |||:::::::   ||||:
    226 ySerTyrTyrLeuLeuLysSerSerGlnAsnSerLysValIleProPheA 243

    719 GCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATC...CGC 765
        ::      |||      ||||||:::|||:::::::::   |||
    243 laProLeuArgAsnLysAspGlySerGlyTyrThrValSerIleSerAla 259

    766 CCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTT 815
        ||||||:::               :::::::::      |||:::   ::
    260 ProValAspPheThrAspLeuGlnAspGluValAlaIleAlaValArgMe 276

    816 CAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGT 865
        :|||:::   :::|||      |||   :::      :::||||||:::
    276 tAsnGlnIleValGluLysGluIleMetLysGlyIleSerGlnTyrMetT 293

    866 TTATGTACAACCGCTATAAAACG 888
        ::::::::::|||:::||||||
    293 rpLeuHisArgArgPheLysThr 300

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18663

seq_documentation_block:
ID   AAW18663 standard; Protein; 387 AA.
XX
AC   AAW18663;
XX
DT   24-JUL-1997  (first entry)
XX
DE   Fragmented human NF-H gene +2 frameshift mutant product.
XX
KW   Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
KW   detection; antibody; probe; cancer; neoplasia; neurodegenerative;
KW   Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
KW   Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
KW   PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
KW   cardiovascular; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1..387
FT                   /note= "X corresponds to a stop codon in the
FT                          accompanying DNA file, AAT69796"
XX
PN   WO9712992-A2.
XX
PD   10-APR-1997.
XX
PF   02-OCT-1996;   96WO-IB01106.
XX
PR   11-JAN-1996;   96US-0009832.
PR   02-OCT-1995;   95GB-0020080.
XX
PA   (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
PA   (UYUT-) UNIV STATE UTRECHT.
XX
PI   Burbach JPH,  Grosveld FG,  Van Leeuwen FW;
XX
DR   WPI; 1997-226235/20.
DR   N-PSDB; AAT69795.
XX
PT   Use of mutant genes having frame:shift mutation(s) - for developing
PT   prods. for the diagnosis, prevention and treatment of associated
PT   diseases, e.g. cancer or neuro:degenerative disease
XX
PS   Claim 22; Fig 9; 123pp; English.
XX
CC   AAW18663 and AAW18664 are +2 and +1 frameshift mutations, respectively,
CC   of a sequence comprising fragments of the coding sequence of the
CC   human neurofilament subunit NF-H gene corresponding to nucleotides
CC   1-1162 of the wild-type NF-H gene. This region contains GAGAG motifs.
CC   Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
CC   microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
CC   and H and amyloid A4 genes are claimed. All these genes share a common
CC   GAGAN motif (N= A, G, C or T), which is the site of common GA
CC   dinucleotide deletion(s) that cause neurodegenerative disorders.
CC   Antigenic peptides used for the production of antibodies, and small
CC   nucleic acid sequences derived from frameshift mutants are used in the
CC   diagnosis, prevention and treatment of cancer and neurodegenerative
CC   diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
CC   syndrome, frontal lobe dementia (Pick's disease), progressive
CC   supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
CC   disease, multiple sclerosis, and other degenerative diseases such as
CC   cardiovascular disease and rheumatoid arthritis.
XX
SQ   Sequence   387 AA;


alignment_scores:
           Quality:  134.00              Length:    311
             Ratio:  1.055                 Gaps:     19
 Percent Similarity:  40.836   Percent Identity: 25.723

alignment_block:
 US-09-303-518D-571 x AAW18663   ..

 Align seg 1/1  to: AAW18663  from: 1  to: 387

       8 GTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTG 57
```

```
    |||   ::::::|||      |||         ||||||::::::|||
125 ValAlaGlnAlaGlyArgAlaProProLeuAlaProProAlaAlaSerTh 141

 58 TTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGC. 106
                      ::::::||||||         |||::::::
141 rArgGlyHisGlyArgPro***AlaProCysProProArgProAlaAlaS 158

107 ............................ACACGCTGGGAAACCGGCTCG 127
                                 |||||||||
158 erValAlaGlnAlaProProGlnAlaProThrArgTrp..... 170

128 GACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCC 177
                                   |||      |||:::
171 ............................ThrArg***AlaThr..... 175

178 AATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGT 227
        ||||||||::
176 .......GlyArgArgAla.............................. 179

228 TTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCG......... 268
                        |||   |||:::   ||||||
180 ...................AlaTrpTrpArgTrpProProHisAlaVal 189

269 CGTTTTTCAAAAAACCGGAAG.....................ACATCGAA 297
    |||   |||:::   |||:::                    ||||||
190 ArgArgSerSerCysArgArg***ThrThrAlaSerProGlyThrSerTh 206

298 ACAATGTTCA......AAGCGGTACACGGCTGGGAACACGTGCAGCAGGC 341
     :::|||:::      :::|||   |||||||:::
206 rArgCysGlySerTrpArgArgThrThrAlaAla..... 217

342 TTTGG...ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCA 388
      |||   :::|||   :::   |||   ||||||||||||      |||
218 ..TrpArgAlaArgLeuArgArgCysGlySerSerArgArg......Ala 231

389 GCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTG 438
    |||   :::||||||::::::|||||||||   |||   |||   :::
232 AlaProLeuTrpAlaSerCysThrSerAlaArgSerAlaArgCysAlaAl 248

439 ACCGCCATGTACA.............................AGCCGCC 458
          |||:::                              ||||||
248 aArgCysCysAlaTrpAlaArgArgAlaValSerTyrAlaTrpSerArgS 265

459 GAAAATCAAAGCGATAGACAAAATCATGC..............AGGCGG 493
           :::|||   ::::::::::|||                 |||
265 erThrCysSerArgThrSerArgThrCysAlaSerAla***ThrThrArg 281

494 GCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTC 543
      |||         :::   :::   ||||||      ::::::::::||
282 ProGlySerGluArgArgProArgArgArgProAlaArgTrpArgAlaSe 298

544 AAACAAATCATCAAGGCCCTGCGCGCGGGCG............AGGCAAC 581
    |::::::      ||||||      ||||||            |||:::
298 rArgArg......ArgProArgArgArgAlaTrpThrCysArgArgArgA 313

582 CATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGT 631
           |||                  ||||||:::||||||:::|||
313 rgArgArgCys.................ArgArgSerAlaAlaThrCys 323

632 GGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAA 681
    |||                                  |||:::::
324 GlyAlaThr..........................ThrArgLysAr 330

682 TTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCT 731
    :|||   :::                  |||   ||||||   :::   :
330 gTrpAlaSer.................Cys...SerAlaArgSerArgA 340

732 GCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAAT 781
    ::|||:::   :::         ||| :::|||   |||
340 laProAlaProArgArgArgArgCys.ArgProArgArgAlaThrPro** 356

782 TGAACGGCAACAAAGCCCACGATGCCGCCGT 812
     :::               ||||||   |||
```

```
     356 *SerAlaThr***ArgArgArgCysAlaArg 366

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814

seq_documentation_block:
ID   AAB59814 standard; Protein; 1022 AA.
XX
AC   AAB59814;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #5.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1022 AA;


alignment_scores:
            Quality:  134.00              Length:    400
              Ratio:  0.859                Gaps:     19
 Percent Similarity:  39.000   Percent Identity: 22.500

alignment_block:
 US-09-303-518D-571 x AAB59814   ..

 Align seg 1/1  to: AAB59814  from: 1  to: 1022

      39 AACCGCCATGCACATCCTGTTGACCGCCCT.........GCTCAAATGCC 79
         ::::::|||:::      |||:::|||             ::::::
      11 SerGlnHisValProAlaValSerArgThrValProHisGlyArgAlaGl 27

      80 TCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAA.......... 119
          |||||||||      :::|||:::|||   ||||||:::
      27 yLeuProAlaGluArgLeuLeuAlaAla...AlaGlyArgArgGlyGlyG 43

     120 ......................CCGGCTCGGACATCTGGCGTTTTACCT 146
                                |||||||:::      ::::::|||
```

```
 43 lyAspGlnIleLeuGluAlaAlaProAlaGluGlnValSerAlaLeuPhe 59

147 TTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTT 196
       :::|||||||||||||
 60 ArgSerGlyArgProArgPro.......................:.. 66

197 TGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA 246
           |||::::::||||||:::::      |||::::::   :
 67 ..........HisValSerGlyGlnGlnHis......GlyGlyAlaValL 78

247 TGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGA 296
    ::||||||            |||:::   |||      :::      |||
 78 euArgPhe............ArgLeuGlnGlnHisArgAlaAlaLeuArg 90

297 AACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGG 346
    |||                  |||   |||   ||||||||||||
 91 Asn..................ArgProGlyArgArgAlaAlaGlyAlaHi 101

347 ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT 396
     |||:::||||||   |||   :::::::::      |||::::::
101 sGlnAlaArgArgLysAlaTyrArgArgSerProGlyArgHisVal.... 116

397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446
                                  ::::::|||   :::||||||
117 ............................AspLeuProLeuGluArgHis 123

447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGG.. 494
          :::   ::::::::::::|||||||||   ::::::   |||
124 GluGlySerArgGlnHisArgGlnAspArgGlnLeuGluGlyAspGlyHi 140

495 .......CAGGGT..................................... 500
            ||||||
140 sArgLeuGlnGlyGlyAspGlnLeuGlyAlaProAlaGlyProAlaValG 157

501 .........................................GCGCGGCAA 509
                                                :::|||
157 lnAspArgArgGlyLysLeuArgAspArgSerGluAlaProGlyArgAla 173

510 AGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGG 559
       :::|||||||   |||      ::::::::::::          |||||
174 AlaArgAsnArgArgHisLeuSerAlaHisSerArgArgAlaLeuGlnGl 190

560 CCCT........................................GCGC 567
    ||||                                        :::|
190 yProGlnGlyArgAspAlaGlyGluIleLeuTyrLeuProAspLeuSerA 207

568 GCGGGCGAGGC.......................AACCATCATCCTGCC 593
    ||   |||:::                      ::::::|||||||||
207 rgAspArgAlaLeuArgGluArgLeuArgProGluGlyArgHisProAla 223

594 ..........................................CGACCACG 601
                                               ::
224 ValAlaValLeuGlnGlyLeuArgArgArgGlnGluIleProAlaAspGl 240

602 TCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTT......TTTC 645
    :|||   ::::::|||      ||||||   |||   |||      ||||
240 uProHisGlyCysGlyGlyThrArgArgAspGlyThrPheGluAspPheA 257

646 GG........................................... 647
    ||
257 rgAlaTrpArgArgGlnValAlaArgLeuProArgAsnLeuProGlyVal 273

648 CAAACCTGCATACAC..............CATGACACTGGCGGCAAAAT 682
    :::      :::|||             |||:::      |||:::
274 GluArgSerValHisProHisArgArgArgHisGlnArgGlnGlyArg.. 289

683 TGGCACACGTCAAAGG............................... 698
     |||      ||||||
290 .GlyArgLeuGlnArgHisAspArgArgHisProArgGlySerGlnAlaA 306

698 ........................................... 698

306 spProHisGlyArgAlaLeuHisArgLeuProLeuPheGlnGluGluPro 322
```

```
699 CGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCT 748
    ||||||:::   ::::::   :::|||:::      ||||||||||||||
323 ArgGluAspAlaAlaLeuGlyPheArgValHisProArgArgThrArgLe 339

749 TCGTGTTGCACATCCGCCCCGTCCA.....................A 774
    |   |||   :::      |||                         |
339 uSerValAspGlnAlaArgArgAspArgHisGlyAlaAspGluGlyIleA 356

775 GGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAA 824
    ||   ::::::   |||::::::   |||   ||||||   :::||||||
356 rgGlnValGlnProGlnArgGlnArgArgHisArgArgGlySerProGln 372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824

seq_documentation_block:
ID AAB59824 standard; Protein; 1605 AA.
XX
AC AAB59824;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #1 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1605 AA;

alignment_scores:
Quality: 134.00 Length: 400
Ratio: 0.859 Gaps: 19
Percent Similarity: 39.000 Percent Identity: 22.500

alignment_block:
US-09-303-518D-571 x AAB59824 ..

Align seg 1/1 to: AAB59824 from: 1 to: 1605

```
 39 AACCGCCATGCACATCCTGTTGACCGCCCT.........GCTCAAATGCC 79
595 SerGlnHisValProAlaValSerArgThrValProHisGlyArgAlaGl 611

 80 TCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAA..........119
611 yLeuProAlaGluArgLeuLeuAlaAla...AlaGlyArgArgGlyGlyG 627

120 .....CCGGCTCGGACATCTGGCGTTTTACCT 146
627 lyAspGlnIleLeuGluAlaAlaProAlaGluGlnValSerAlaLeuPhe 643

147 TTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTT 196
644 ArgSerGlyArgProArgPro..... 650

197 TGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA 246
651 .....HisValSerGlyGlnGlnHis......GlyGlyAlaValL 662

247 TGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGA 296
662 euArgPhe.....ArgLeuGlnGlnHisArgAlaAlaLeuArg 674

297 AACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGG 346
675 Asn.....ArgProGlyArgArgAlaAlaGlyAlaHi 685

347 ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT 396
685 sGlnAlaArgArgLysAlaTyrArgArgSerProGlyArgHisVal.... 700

397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446
701 .....AspLeuProLeuGluArgHis 707

447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGG.. 494
708 GluGlySerArgGlnHisArgGlnAspArgGlnLeuGluGlyAspGlyHi 724

495 .......CAGGGT..... 500
724 sArgLeuGlnGlyGlyAspGlnLeuGlyAlaProAlaGlyProAlaValG 741

501 .....GCGCGGCAA 509
741 lnAspArgArgGlyLysLeuArgAspArgSerGluAlaProGlyArgAla 757

510 AGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGG 559
758 AlaArgAsnArgArgHisLeuSerAlaHisSerArgArgAlaLeuGlnGl 774

560 CCCT.....GCGC 567
774 yProGlnGlyArgAspAlaGlyGluIleLeuTyrLeuProAspLeuSerA 791

568 GCGGGCGAGGC.....AACCATCATCCTGCC 593
791 rgAspArgAlaLeuArgGluArgLeuArgProGluGlyArgHisProAla 807

594 .....CGACCACG 601
808 ValAlaValLeuGlnGlyLeuArgArgArgGlnGluIleProAlaAspGl 824

602 TCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTT......TTTC 645
824 uProHisGlyCysGlyGlyThrArgArgAspGlyThrPheGluAspPheA 841

646 GG..... 647
841 rgAlaTrpArgArgGlnValAlaArgLeuProArgAsnLeuProGlyVal 857

648 CAAACCTGCATACAC.....CATGACACTGGCGGCAAAAT 682
858 GluArgSerValHisProHisArgArgArgHisGlnArgGlnGlyArg.. 873

683 TGGCACACGTCAAAGG..... 698
874 .GlyArgLeuGlnArgHisAspArgArgHisProArgGlySerGlnAlaA 890

698 ..... 698
890 spProHisGlyArgAlaLeuHisArgLeuProLeuPheGlnGluGluPro 906

699 CGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCT 748
907 ArgGluAspAlaAlaLeuGlyPheArgValHisProArgArgThrArgLe 923

749 TCGTGTTGCACATCCGCCCCGTCCA.....A 774
923 uSerValAspGlnAlaArgArgAspArgHisGlyAlaAspGluGlyIleA 940

775 GGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAA 824
940 rgGlnValGlnProGlnArgGlnArgArgHisArgArgGlySerProGln 956
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:

ID AAB59817 standard; Protein; 999 AA.
XX
AC AAB59817;
XX
DT 04-APR-2001 (first entry)
XX
DE TutD protein #8.
XX
KW Toluene degradation; enzyme; waste degradation; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23625, AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 5; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ Sequence 999 AA;

```
alignment_scores:
           Quality:  132.50              Length:    290
             Ratio:  1.162                Gaps:     13
 Percent Similarity:  39.310   Percent Identity: 27.241

alignment_block:
 US-09-303-518D-571/rev x AAB59817   ..

 Align seg 1/1  to: AAB59817  from: 1  to: 999

     814 ACACGGCGGCATCGTGGGCTTTGTTGCCGTTCAATTCCCCTTGGACGGGG 765
         |||||||||   |||:::   ||||||||||:::            |||||
      75 ThrArgArgLysArgAsnCysCysCysArgAlaAlaArgArgPheArgGl 91

     764 C.................GGATGTGCAACACGAAGCCTTGTCCGTCGGG 733
         |                 :::||||||   |||          ||||||
      91 yProLeuArgAlaAlaProAlaCysAlaArgArgPheAlaTyrArgArgL 108

     732 CAGGCGTTCGCAGCAGAAAAACAGGGTTTTCACGCCTTTGACGTGTGCCA 683
           |||   |||      :::   |||   ||||||
     108 euGlyProArgCysThrSerArgGlyArgSerArgCys..... 120

     682 ATTTTGCCGCCAGTGTCATGGTGTATGCAGGTTTGCCGAAAAAATCCGCC 633
                       :::   |||   :::      |||   :::   ||
     121 .........SerProAspArgCysCysArgTrpSerArgCysSerSerPr 134

     632 CACACGCCGCCGCCTTCCTGCGGAGAAGGGACGTGGTCGGGCAGGATGAT 583
         |:::   ||||||   |||   ::::::      |||      |||
     134 oSerProArgArgCysProProSerSerProAlaGlyAlaProGlyAlaT 151

     582 GGTTG.....................CCTCGCCCGCGCGCAGGGCCTTGA 554
                                     |||      |||||||||
     151 hrCysSerArgArgProPheSerArgSerArgAspSerAlaGlyPro... 166

     553 TGATTTGTTTGACCCCTTGTATGCCGGTGGGCGCGGTTTTGCCTTTGC.. 506
                               |||   |||||||||      |||
     167 .......................ArgAlaAlaArgPheArgArgCysAr 175

     505 ................CGCGCACCCTGCCCG..............CCTG 487
                         ||||||   ||||||               |||:
     175 gAspAlaCysGluArgArgAlaArgCysProGlyProArgSerAlaProS 192

     486 CATGATTTTGTCTATCGCTTTGATTTTCGGCGGCTTGTACATGGCGGTCA 437
         ::            |||        |||   |||      ||||||
     192 erIleArgArgGlySerArgAspArgSerArgAlaSerArgSerArgSer 208

     436 GGTGGAACGGAAGCTGCTGGCTGATGTAGCGTCCGCCCAAATCGTAGCTG 387
           |||:::      :::|||      ::::::         |||
     209 ArgGlySerProLeuCysGlyAlaThrAlaThrSerCysProArgArgAr 225

     386 CCGATGTGCGGCGTGATGAACAGCAGCCCTTCGCCCTTGTCCAAAGCCTG 337
         |||||||:::            :::|||         ||||||   |||
     225 gArgCysSer.........IleGlyAlaSerSerGlyCysProHisProP 239

     336 CTGCACGTGTTCCCAGCCGTGTACCGCTTTGAACATTG............ 299
           :::|||         |||   |||:::  :::
     239 roValArg.........ArgSerProValAsnSerSerLysArgAlaHis 252

     298 TTTCGATGTCTTCCGGTTTTTTGAAAAACGCGGGGGCAAGTTCCA..... 254
           ||||||                :::   ||||||   :::
     253 ArgArgCysThrAlaArgArgGlyArgPheArgGlyProThrSerArgAs 269

     253 .........AACCGCATTTTGCCGTTTCCGCAAAAACGGCTTTGACCGT 214
                 :::|||         :::|||:::            |||
     269 pThrGlyArgArgArgCysTrpArgTrpProArg...........ProA 282

     213 CTGCGTGTCGG............................GGTTCAAAC 194
           ||||||                              :::|||:::
     282 rgArgCysArgCysSerArgArgTrpGlyArgProLeuTrpAlaSerGly 298

     193 CCGCCTGCCGCATATTGGCGACGATGCGCGCGCGGTCTTCCTTTAAAAGG 144

           |||   |||   |||||||||   :::
     299 CysProArgAlaArgTrpArgArgGlySer..................... 308

     143 TAAAACGCCAGATGTCCGAGCCGGTTTCCCAGCGTGTGCAGACAGGAAAG 94
                   :::      :::|||      ::::::|||:::   |||
     309 .........AsnTrpSerSerGlyArgSerSerAlaAlaSerProLysA 322

      93 CGACAGCAGGGAGAGGCATT 74
           |||:::||||||   :::
     322 rgThrCysGlyArgArgVal 328

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU40508

seq_documentation_block:
ID   AAU40508 standard; Protein; 354 AA.
XX
AC   AAU40508;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #1404.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59512.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 1703; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence   354 AA;


alignment_scores:
           Quality:  132.00            Length:    222
             Ratio:  1.307               Gaps:      9
 Percent Similarity:  45.495   Percent Identity: 27.027

alignment_block:
 US-09-303-518D-571 x AAU40508   ..

 Align seg 1/1  to: AAU40508  from: 1  to: 354

     203 CCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGT 252
         |||   |||||||||      |||   |||||||||      :::
      34 ProProArgArgArg......ProGlnLeuArgLysProGlyArgHisPr 48

     253 TTGGAACTTGCCCCCGCGT..............................  271
         ::::::   |||||||||
      48 oTyrArgSerProProArgGlnArgProArgProLeuTrpProAspCysA 65

     272 .....TTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAA......AG 310
              :::|||   :::||| :::|||   ::::::||||||      ||
      65 laAlaTyrSerGlySerArg.AsnHisProArgGlyValGlnGlyTrpSe 81

     311 CGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGG 360
         ||||   |||:::               |||:::   :::
      81 rGlyGlyArgIleHisArgTrpProArgGlyLeuProArg.......... 94

     361 CTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTA 410
              ||||||   |||   :::|||:::   |||   |||      |||
      95 .....ValHisProAlaHisGlnArgArgSerArgThrGlySerArgLeu 109

     411 CATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGA 460
                                                         ::
     110 ..................................................As 110

     461 AAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAA 510
         ::::   :::   |||   :::|||      |||:::   :::|||   |
     110 pArgAlaThrHisArgAlaHisHisArgGlnGlyArgProSerArgProA 127

     511 GGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGC 560
         ||   :::|||      |||   :::               |||:::
     127 rgProHisArgArgAlaArgProSerGlyTrpLeuGlnProHisArgArg 143

     561 CCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTC 610
              :::||||||   :::   ||||||::::::|||          ::
     144 ArgProGlnGlyArgArgGlyGlyHisProThrHisProSerAlaAlaGl 160

     611 CGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATAC 660
         :      |||   ||||||   |||:::      |||:::        |
     160 yArgProArgGlyArgArgProGlyAsn......ArgArgLeuHisHisH 175

     661 ACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCT 710
         |||||      :::|||   :::   :::::::::   |||:::
     175 isHisTrpCysSerGlyHisLeuHisValAsnArgProArgAsn...... 189

     711 GTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACA 760
                   |||      ::::::|||   :::||||||      ||
     190 .........CysArgArgGlnSerHisArgGlyAsnLeuArgGlnHisHi 203

     761 TCCG.................CCCCGTCCAAGGGGAATTGAACGGCAAC 792
         ||||                 |||||||||::::::   :::|||
     203 sProAlaHisArgArgCysHisProArgProGlnAlaHisHisArgProA 220

     793 AAAGCCCACGATGC 806
                 |||
     220 laLeuAlaValCys 224

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29192

seq_documentation_block:

ID   AAY29192 standard; Protein; 262 AA.
XX
AC   AAY29192;
XX
DT   25-OCT-1999  (first entry)
XX
DE   Amino acid sequence of a virulence factor encoded by ORF25892c.
XX
KW   Human pathogen; virulence polypeptide; virulence factor;
KW   pathogenic infection; Pseudomonas aeruginosa infection.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO9927129-A1.
XX
PD   03-JUN-1999.
XX
PF   25-NOV-1998;   98WO-US25247.
XX
PR   25-NOV-1997;   97US-0066517.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Ausubel F,  Cao H,  Drenkard E,  Goodman HM,  Mahajan-Miklos S;
PI   Rahme LG,  Tan M,  Tsongalis J;
XX
DR   WPI; 1999-357851/30.
XX
PT   Virulence factors useful in developing disease treatments
XX
PS   Disclosure; Fig 3; 228pp; English.
XX
CC   The present sequence represents a Pseudomonas aeruginosa polypeptide
CC   sequence. P. aeruginosa is an opportunistic human pathogen present in
CC   soil water and plants. The specification describes virulence polypeptides
CC   and nucleic acid sequence encoding such polypeptides. These sequences
CC   can be used to identify a compound which is capable of decreasing the
CC   expression of a pathogenic virulence factor. Compounds that inhibit
CC   the expression or activity of virulence factor polypeptides can be
CC   used to treat pathogenic infections, especially where the infection
CC   is a P. aeruginosa infection.
CC   note: the sequences given in the specification were poorly legible, and
CC   in some instances assumptions were made as to the identity of the
CC   residue; it is therefore possible that the sequence given below is
CC   not entirely correct.
XX
SQ   Sequence   262 AA;


alignment_scores:
           Quality:  131.50            Length:    253
             Ratio:  1.096               Gaps:     14
 Percent Similarity:  47.431   Percent Identity: 30.040

alignment_block:
 US-09-303-518D-571 x AAY29192   ..

 Align seg 1/1  to: AAY29192  from: 1  to: 262

     144 CCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGG 193
         |||   :::|||||||||          :::|||          ||||||
       4 ProAlaGluGlyArgProValHisLeuGlnArgAlaGlyIleAlaGly.. 19

     194 GTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCA 243
                              ||||||:::::::::   :::   ::::::
      20 ................ArgAspGlyArgGluHisGlnAlaArgGlnAlaA 31

     244 AAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACAT 293
              |||   |||          ||||||   |||   :::|||:::
      31 laAlaArgProGly.........ArgValAlaGlnGlyAlaGlyGluCys 44

     294 CGAAACAATGTT.........CAAAGCGGTACACGGCTGGGAACACGTGC 334
         |||::::::             :::|||||||   |||          |||::
```

```
 45 ArgGlyGlnProAlaTrpArgArgSerGlyGluArgProGluArgArgGl 61

335 AGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATC 384
    :::::::   |||   |||:::                |||      |||
 61 yValAlaProGlyAlaGlyGln............ValHisArgArgHisA 74

385 GGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCC...... 428
      |||   |||   ||||||      :::          |||
 74 spGlnAlaArgArgGlyArgGlnAlaArgProGlyAspArgSerArgArg 90

429 ...GTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAG 475
          ||||||||||:::         ||||||         :::|||::
 91 ArgAspProProAspHisProGlyProAlaAla.........AlaAspGl 104

476 ACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCC 525
    :::::::   ::::::   :::   :::|||:::||||||:::|||   |
104 nGluGlnProGlyAlaAspArgArgThrArgArgArgGlnAspArg...H 120

526 ACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGA 575
    ||||||::      |||   ::::::||||||      :::   ||||||
120 isArgArgGlyProGlyProAlaHisHisGlnArgArgSerAlaGlyArg 136

576 ...GGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCG 622
          ::::::   ||||||   |||         :::|||   |||||
137 ProGlnGlyGlnAlaProAlaGlyProGlyHisGlyGlyAlaAspArgAr 153

623 G............CGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATAC 660
    |           |||||||||   |||
153 gCysGlnValProArgArgValArgGly..................... 162

661 ACCATGACACTGGCGGC......AAAATTGGCACACGTCAAAGGCGTGAA 704
              ||||||      ::   |||         |||
163 ..ThrProGluGlyGlyProGlnArgThrGlyGlnAlaGlyArgProGly 178

705 AACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGT 754
    :::||||||         |||      |||::::::   |||   |||
179 HisProValHis......ArgArgThrAlaHisHisGlyArgArgArgGl 193

755 TGCA.....................CATCCGCCCCGTCCAAGGGGAA 780
     :::                        |||              |||
193 nGlyGlyArgCysHisGlyArgArgGlnHisAlaGlnAlaGlySerGlyA 210

781 TTGAACGGCAACAAAGCCCACGATGC.........CGCCGTGTTCAACCG 821
      :::|||           ||||||         |||||||||:::|||
210 laArgArgAlaAlaLeuArgArgCysTyrTyrProArgArgValSerPro 226

822 CAATACCGA 830
      :::|||
227 ValHisArg 229
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29184

seq_documentation_block:

ID AAY29184 standard; Protein; 558 AA.
XX
AC AAY29184;
XX
DT 25-OCT-1999 (first entry)
XX
DE Amino acid sequence of a virulence factor encoded by ORF25103c.
XX
KW Human pathogen; virulence polypeptide; virulence factor;
KW pathogenic infection; Pseudomonas aeruginosa infection.
XX
OS Pseudomonas aeruginosa.
XX
PN WO9927129-A1.
XX
PD 03-JUN-1999.
XX
PF 25-NOV-1998; 98WO-US25247.
XX
PR 25-NOV-1997; 97US-0066517.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
PI Rahme LG, Tan M, Tsongalis J;
XX
DR WPI; 1999-357851/30.
XX
PT Virulence factors useful in developing disease treatments
XX
PS Disclosure; Fig 4; 228pp; English.
XX
CC The present sequence represents a Pseudomonas aeruginosa polypeptide
CC sequence. P. aeruginosa is an opportunistic human pathogen present in
CC soil water and plants. The specification describes virulence polypeptides
CC and nucleic acid sequence encoding such polypeptides. These sequences
CC can be used to identify a compound which is capable of decreasing the
CC expression of a pathogenic virulence factor. Compounds that inhibit
CC the expression or activity of virulence factor polypeptides can be
CC used to treat pathogenic infections, especially where the infection
CC is a P. aeruginosa infection.
CC note: the sequences given in the specification were poorly legible, and
CC in some instances assumptions were made as to the identity of the
CC residue; it is therefore possible that the sequence given below is
CC not entirely correct.
XX
SQ Sequence 558 AA;

alignment_scores:

```
            Quality:  131.50              Length:    339
              Ratio:  0.913                Gaps:     16
 Percent Similarity:  42.478   Percent Identity: 25.074
```

alignment_block:

US-09-303-518D-571 x AAY29184 ..

Align seg 1/1 to: AAY29184 from: 1 to: 558

```
 18 CAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA.............CA 52
    |||||||||:::|||:::   :::|||   |||              ||
 17 GlnAlaValAlaProLeuHisHisArgSerAlaThrAlaGlyGlnGlyHi 33

 53 TCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGT 102
    |      ||||||   ::::::::::|||   |||            |||:
 33 sArgProAspArgArgGlyArgGlnProHisProHisGlyAspArgLeuG 50

103 CTGCA.....................CACGCTGGGAAACCGGCTCGGACA 131
    ::|||                    :::      |||||||:::
 50 lnAlaGlyGlyThrGlySerSerArgProSerProAspProAlaGluAsp 66

132 TCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAA.. 179
       :::            :::||||||   |||:::|||::::::
 67 ArgAlaArgGlyAlaGluGluGlyArgArgArgSerHisGlnGluAlaPr 83

180 ....TATGCGGCAGGCGGGTTTGAACCCCGACACGCA............. 212
          |||:::||||||::::::      :::|||
 83 oGlyGlnAlaGlyGlyGlyTyrArgGlnAlaArgAlaArgIleArgArgP 100

213 ......................GACGGTCAAAGCCGTTTTTGCGGA... 236
                           :::|||::::::         |||
100 roArgGlyAspLeuGluValArgGluGlyArgGlyAlaGlyLeuGlyAla 116

236 ..................................................... 236

117 AspProAlaGluAspArgAlaGlyGlnAlaGlyAspGlyGlyGlyAlaAl 133

237 .AACGGCAAAATGCGGTTT.............................. 254
     :::|||:::   |||
133 aGlnGlyArgProArgGluHisGlyAlaHisProValProAspHisProG 150

255 .....GGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAAC 299
        ||||||   |||         :::   :::   :::   |||:::
```

```
     150 lyProGlyThrGlnProAlaAspGlyArgProAlaArgGlnAspArgGlu 166

     300 AATGTT......CAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTT 343
         |||      |||:::|||   |||   |||:::|||   :::|||||
     167 ProValAlaAlaGlnGlnGlyAspArgArgGlyAsnArgArgSerGlyPh 183

     344 T.................................GGACAAGGGCGAA 357
         |                                 |||:::::|||:
     183 eGlnValAspArgTyrProGlyValGluAspAlaArgGlyArgAlaArgG 200

     358 GGGCTGCTGTTCATCACGCC............................. 377
         ::||||||:::|||   |||
     200 luAlaAlaAlaHisGlyAlaGlyAlaAlaSerAlaSerAspArgProGly 216

     378 ..................GCACATCGGCAGCTACGATTTGGGCGG..... 404
                           ::::::|||:::      |||:::|||
     217 ArgGlyGlySerArgArgValGlnArgArgAlaProPheAlaArgArgPr 233

     405 .ACGCTACATCAGCCAGCAGCTTCCGTTCCACCT................ 437
                 :::||||||      :::   |||
     233 oArgArgSerGluProAlaGluArgLeuValProLeuProArgProAspA 250

     438 ...........GACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGA 476
                 ||||||   ||||||:::|||   :::         |||
     250 rgGlyGlyGlnAspArgValValGlnGlyAlaGlyArgValProLeuArg 266

     477 CAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCA 526
             :::   ||||||   ||||||   :::|||      |||:::||
     267 ......TyrArgGlyGlyAlaGlyAla...AspArgTyrValArgValHi 280

     527 CCGGCATACAAGGGGT.........CAAACAAATCATCAAGGCCCTGCGC 567
         |   :::|||   |||         :::   :::   :::|||      |
     280 sGlyGluThrLeuGlyGlyProProAspArgArgAlaSerGlyLeuArgA 297

     568 GCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCC............ 605
         ||   ||||||::::::   |||   |||      |||
     297 rgLeuArgGlyArgArgLeuProAspArgGlyAspProProGlnAlaLeu 313

     606 ......TTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCA 649
               :::|||||||||            ||||||        ::
     314 LeuGlyGlyAlaAlaGlyArg............GlyGlyGluGlyProSe 326

     650 AACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGC 699
         :   ||||||:::|||:::      |||      |||           |
     326 rGlyCysIleGlnHisSerProProGlyAlaArgGlyArgThrProAspA 343

     700 GTGAAAACCCTGTTTTT 716
         ||:::::|||:::::
     343 rgGlnSerProAlaTyr 348

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23029

seq_documentation_block:
ID   ABG23029 standard; Protein; 3640 AA.
XX
AC   ABG23029;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23020.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87216.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 53388; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   3640 AA;


alignment_scores:
            Quality:  131.50              Length:    324
              Ratio:  0.946                Gaps:     19
 Percent Similarity:  42.901   Percent Identity: 26.852

alignment_block:
 US-09-303-518D-571 x ABG23029   ..

 Align seg 1/1  to: ABG23029  from: 1  to: 3640

      63 CGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGC 112
         |||||||||   :::      |||||||:::|||   ||||||
    1858 ArgProAlaGlyVal...ArgProAlaAlaAlaAlaLeuSer........ 1870

     113 TGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGC 162
               |||:::         |||||||||   ::::::|||:::|||
    1871 .......ProGlyProLeuArgGlyValLeuCysLeuGlnGlyGlnPro. 1884

     163 GCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCA 212
             :::   |||:::   ||||||   ||||||:::
    1885 ..LeuGlnLeuArgSerArgAlaAlaProGlyPheGlnAlaThrLeuArg 1900

     213 GACGGTCAAAGCCGT.......................TTTTGCGGAAA 238
             |||   :::|||                        |||||||::
    1901 GlyGlyProGlyArgGlyGlyAlaGlyProAlaGlySerArgCysGlyAr 1917

     239 CGGCAAAATGCGGTTTGGAACTTGCCCC......CGCGTTTTTCAAAAAA 282
         :   ::::::   ||||||      |||               |||::::
    1917 gProGlnGlnValPheGlyHisHisProProArgAlaGlnArgGlnArgA 1934

     283 CCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGT 332
         ::|||   |||         :::::::::            :::   |||
```

```
1934 snGlyAlaHisSerLeuAlaAlaArgAlaHisArg***CysAlaProArg 1950

 333 GCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCAT.......... 371
     ||||||      :::   ||||||      |||   |||
1951 AlaAla......AlaAlaGlyArgSerProAlaArgHisArgValLeuVa 1965

 372 ..........CACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTAC 411
              |||   |||:::|||                         |
1965 lGlyProGlyHisArgAlaGluArgAlaProSerProAspThrGlnGluH 1982

 412 ATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAA 461
     ||:::      ::::::||||||         |||      ::::::
1982 isHisGlyAspSerGlyValProGluGlyProHis......CysGly*** 1996

 462 AATCAAAGCGATAGA............CAAAATCATGCAGGCGGGCAGGG 499
     :::   :::||||||                |||::::::   |||||
1997 HisProAlaAspArgCysCysAlaGlyProValHisGlyAlaGlnGlnGl 2013

 500 TGCGCGGCAA................................AGGCAAA 516
     |||||||:::                                  :
2013 yAlaArgMetProLeuValProGluAlaAspAlaAlaGlnAlaGlyGlyH 2030

 517 ACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCG 566
     ::   ||||||         |||   :::   :::|||:::
2030 isAspAlaHisProAlaGlyArgAspHisArgGlyHisArgAspAlaHis 2046

 567 CGCGGGCGAGGCAACCAT......CATCCTGCCCGACCACGTCCC..... 605
     |||   |||   :::|||      |||      ||||||:::|||
2047 ArgHisArgArgGlnHisProGlnHisHisArgArgProHisProProGl 2063

 606 ............................TTCTCCGCAGGAAGGCGGCGGC 627
                                :::   ||||||   |||
2063 yGlnLeuGlyArgAlaGlyThrThrAlaLeuArgAlaGlySerArgValT 2080

 628 GTGTGGGC..............GGATTTTTTCGGCAAACCTGCATACAC 662
       ::::::             |||:::      :::   |||   |||
2080 hrIleSerAspGlyGlyValProGlyLeuGlnProAspLeuCysProHis 2096

 663 ....................................CATGACACTGG 673
                                         |||:::||
2097 AlaHisProHisAlaLeuProArgAlaGlnArgGlyAlaProAspAlaGl 2113

 674 CGGCAAAATTGGCACACGTCAAAGGCGTGAA..............AACC 708
     ||||:::   |||   |||        :::             :::|
2113 yGlyArg...GlyAspArgGlyProGlyGlnAlaLeuGlyProAlaGluP 2129

 709 CTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCA 758
     ||:::   |||   |||   |||      |||:::   |||   :::
2129 roAlaValLeuTrpArgArgPro......ArgAlaTrpLeuProLeuLeu 2143

 759 CATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCG 808
     |||         |||||||||:::::|||      :::
2144 His.........ProArgGlyPheGlnArgGlyProGly.......... 2153

 809 CCGTGTTCAACCGCAATACCGA 830
           |||||||||   |||
2154 .......GlnProGln***Arg 2158
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03914

seq_documentation_block:

```
ID   ABG03914 standard; Protein; 560 AA.
XX
AC   ABG03914;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #3905.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS68101.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 34273; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   560 AA;
```

alignment_scores:

```
           Quality:  130.50              Length:    369
             Ratio:  0.837                Gaps:     19
 Percent Similarity:  42.276   Percent Identity: 24.932
```

alignment_block:

```
US-09-303-518D-571 x ABG03914   ..

Align seg 1/1  to: ABG03914  from: 1  to: 560

  27 TCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAAT 76
     |||||||||   :::      ::::::   ::::::||||||   :::
 205 SerProPheTyrArgSerSerSerArgGluThrSerArgProProGluGl 221

  77 GCCT...CTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT...GGGAAAC 120
       |||      |||   :::      |||||||||   |||   |||:::|
 221 yProArgArgProArgAlaProAlaLeuSerAlaProAlaProGlyGlnP 238

 121 CGGCTCGGACATCTGGCGTTTTACCT.................TTTAAA 152
     ||||||||               |||                  |||
 238 roAlaArgProArgProArgGluProValProCysGlyAlaValPheThr 254

 153 GGAAGACCGCGC.....................GCGCAT........... 170
     :::|||   |||                     :::|||
```

```
255 AlaArgAspArgLeuArgProProAlaAlaThrSerHisAlaProPheSe 271

171 .............CGTCGCCAATATGCGGCAGGCGGG............. 194
                 ||||||   :::      ||||||
271 rAlaAlaAsnProArgArg***HisArgProGlyGlyProGlyAlaArgA 288

195 .................TTTGAACCCCGACACGCAGACGGTCAAAGCCGT 227
                     ::::::   |||   ::::::|||      |||
288 rgLeuGlyAspAlaGlnLeuSerArgArgSerThrSerGlyAlaProArg 304

228 TTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGC.............. 263
      |||::::::   :::   |||      ||||||
305 ...CysSerGlnThrArgSerArg***ProThrCysValCysValAlaLe 320

264 .CCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCG 312
     |||            :::::::::   ||||||:::      |||:::|
320 uProLeuProProGlyArgSerAlaProHisArgSerCysSerGlnAlaG 337

313 GTACACGGCTG.........GGAACACGT..................... 332
    ||   :::|||         |||   |||
337 lyArgGluLeuProGlyGlnGlyProArgTyrTyrArgHisLeuProGln 353

333 ..............GCAGCAGGCTTTGGACAAGGG.............. 353
                  :::   |||   |||||||||
354 LeuSerIleLeuHisSerIleGlyGluGlyGlnGlyCysGlyPheTrpSe 370

354 ............................CGAAGGGCTGCTGTTCATCACG 375
                                |||   |||::::::   :
370 rGluArgSerPheLysGlyTyrProGluArgProAlaGlyAlaAlaGlyV 387

376 CCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCT 425
    ::::::::   |||           |||           |||
387 alCysArgLeuGlnGlyCysGlyArgArgGlyArgGlyAlaProPheArg 403

426 TCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAG 475
    ::::::      :::   :::   ::::::|||   :::::::::|||::
404 ThrThrAspPheSerSerArgProArgGlyAlaAlaGluArgAlaAspGl 420

476 ACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCC 525
    :      :::||||||:::         |||      ::::::
420 nGlyProArgAlaGlySerProTrpProArgThrThrSerGlyAla.... 435

526 ACCGGCATACAAGGGGTCAAACAAATCATCAAGGC...CCTGCGCGCGGG 572
                ||||||:::::::::   ||||||      ||||||
436 ........GlnArgGlyArgAlaGlnGlyGlnGlyHisThrAlaArgArg 449

573 CGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCG 622
    |||||||||   :::|||:::      |||      :::|||:::|||||
450 ArgGlyAsnSerAsnProGlyProSerArgAlaArgGlnAlaSerArgAr 466

623 GCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTG 672
    |||||||   :::                                 :::|
466 gArgArgProAla.............................ThrSerG 473

673 GCGGCAAAATTGGCACA............CGTCAAAGGCGTGAAAACCCT 710
    ||        |||:::            |||   ||||||::::::|||
473 lyProProArgGlySerProArgProAspArgProArgArgArgSerPro 489

711 GTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACA 760
       :::      |||:::   :::|||:::|||
490 PheTyr......ArgSerSerSerArgGluThrSerArgProProGluGl 504

761 TCCGCCCCGTCCAAGG.........................GGAATTG 783
     |||   |||||||||                         |||
504 yProArgArgProArgAlaProAlaLeuSerAlaProAlaProGlyGlnP 521

784 AACGGCAACAAAGCCCACGA............TGC.............. 806
       :::      ||||||      |||
521 roAlaArgProArgProArgGluProValProCysGlyAlaValPheThr 537

807 ......CGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTC 850
           |||::::::|||                  :::|||||||||
538 AlaArgAspArgLeuArgProProAlaAlaThrSerHisAlaProPheSe 554
```

```
     851 CGACGCA 857
         |   |||
     554 rAlaAla 556
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:

```
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;
```

alignment_scores:

```
            Quality:  130.00              Length:    373
              Ratio:  0.839                Gaps:     18
 Percent Similarity:  41.555   Percent Identity: 21.984
```

alignment_block:

```
 US-09-303-518D-571 x AAB59817   ..

 Align seg 1/1  to: AAB59817  from: 1  to: 999

      77 GCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTC 126
         ||||||   |||   |||            :::|||||||:::   |||||
     296 AlaSerGlyCysProArg............AlaArgTrpArgArgGlySe 308

     127 GGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC 176
         |:::         ::::::      |||:::::|||   :::   :::
     308 rAsn.........TrpSerSerGlyArgSerSerAlaAlaSerProLysA 322
```

```
177 CAATATGCGGCAGGCGGGTTTGAACCCCGACA..................... 208
      ||||||||||||||||   :::   |||
322 rgThrCysGlyArgArgValArgSerAspThrSerAlaArgArgSerArg 338

209 .....................................CGCAGACGGTC 219
                                         ||||||
339 CysProAlaSerSerProIleArgTrpThrGlyArgCysArgArgTrpAr 355

220 AAAGCCGTTT..................................... 229
     :::|||:::
355 gArgProLeuGlyCysSerProArgAlaThrCysThrAlaArgCysGlyA 372

230 ...............................TTGCGGAAACGGCAAAAT 247
                                    |||::::::      :::
372 rgAspGlyCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuArg 388

248 GCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAA 297
    :::   |||      |||   |||      ::::::|||:::|||:::::
389 GlyProTrpAlaAlaProPheArgAlaHisArgSerArgSerThrThrAr 405

298 ACAATGT............TCAAAGCGGTACACGGCTGGGAACACGTGCA 335
    ::::|||           |||:::|||:::           :::
405 gArgCysAlaValArgGlySerSerArgHisAspArgThrAlaSerThrA 422

336 GCAGGCTTTGGACAAGGGCGAAGGGCTGC..................... 364
      |||               |||||||||
422 rgArgProHisLysProProLysGlyCysAlaThrAspIleHisSerGly 438

365 ......TGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGC 408
          |||      |||   :::|||:::   :::      |||:::::
439 ArgTyrCysTrpProArgThrAlaSerSerArgAla......AlaSerGl 453

409 TACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCC 458
    ::::|||||:::      |||                |||      |||
453 yAlaSerAlaLysArgThrArgLeuArgArgArgSerCysProValArgS 470

459 GAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCA 508
           :::|||   |||:::::::::   :::::::::|||::::::
470 erProArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySer 486

509 AAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAG 558
    :::::::::   ||||||   :::   :::      :::         ||
487 SerSerArg...ArgProSerSerGlyArgProTrpSerValProIleAr 502

559 GCCC.........TGCGCGCGGGCG.........AGGCAACCATCATCCT 590
    ||||         |||:::||||||          :::||||||||||
502 gProSerSerIleCysGlyArgAlaValGlyLeuThrSerProSerSerP 519

591 GCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCG.........TGT 631
        :::      :::   ||||||:::|||   |||         |||
519 roLeuAsnArgProPheAlaArgArgSerAlaProAlaSerThrProCys 535

632 GGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAA 681
      |||                                |||   |||:::::
536 ArgArg..............................HisAsnArgArgAr 542

682 TTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCG....... 724
    ::::   :::   :::      :::      |||:::::::::
542 gTyrGlySerArgArgProPheArgArgArgPheAlaCysSerTrpSerS 559

724 ............................................... 724

559 erGlnHisAspProAlaSerGlnAspProGlnArgGlyThrCysProLeu 575

725 ...AACGCCTGCCCG......ACGGACAAGGCTTCG.............. 751
       ||||||||||||      :::   :::||||||
576 ArgAsnAlaCysProGlyTrpAlaProArgAlaSerArgProHisLeuPr 592

752 .............TGT.........TGCACATCCGCCCCGTCCAAGGGGA 779
                |||         |||              |||   ::::
592 oLeuArgArgArgCysLysGluArgCysProPheArgCysSerProAlaS 609

780 ATTGAACGGCAACAAAGCCCACGATGCCG..............CCGTGT 814
    ::      ::::::::::||||||   |||               ||||||
609 erProProThrAlaSerProThrTrpProAlaSerSerGlySerProCys 625

815 .................TCAACCGCAATACCGAATATTGGATACGCCGT 846
                        :::   |||      :::   :::::::::
626 GlyAlaSerGlyAlaSerThrProAlaSerAlaSerTrpAlaHisSerAr 642

847 TTTCCGACGCAGTATCTGT 865
      ||||||   |||   |||
642 gPheArgSerSerThrCys 648
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

ID AAB59827 standard; Protein; 1592 AA.
XX
AC AAB59827;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #4 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1592 AA;

alignment_scores:

Quality:	130.00	Length:	373
Ratio:	0.839	Gaps:	18
Percent Similarity:	41.555	Percent Identity:	21.984

alignment_block:

US-09-303-518D-571 x AAB59827 ..

Align seg 1/1 to: AAB59827 from: 1 to: 1592

```
 77 GCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTC 126
    ||||||   |||   |||           :::||||||:::   |||||
889 AlaSerGlyCysProArg...........AlaArgTrpArgArgGlySe 901

127 GGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC 176
    |:::         ::::::      |||::::::|||   :::   :::
901 rAsn.........TrpSerSerGlyArgSerSerAlaAlaSerProLysA 915

177 CAATATGCGGCAGGCGGGTTTGAACCCCGACA..................... 208
        ||||||||||||||   :::   |||
915 rgThrCysGlyArgArgValArgSerAspThrSerAlaArgArgSerArg 931

209 .....................................CGCAGACGGTC 219
                                        ||||||
932 CysProAlaSerSerProIleArgTrpThrGlyArgCysArgArgTrpAr 948

220 AAAGCCGTTT....................................... 229
     :::|||:::
948 gArgProLeuGlyCysSerProArgAlaThrCysThrAlaArgCysGlyA 965

230 ...............................TTGCGGAAACGGCAAAAT 247
                                  |||::::::      :::
965 rgAspGlyCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuArg 981

248 GCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAA 297
    :::   |||      |||   |||      ::::::|||:::|||:::::
982 GlyProTrpAlaAlaProPheArgAlaHisArgSerArgSerThrThrAr 998

298 ACAATGT............TCAAAGCGGTACACGGCTGGGAACACGTGCA 335
    :::::||           |||:::|||:::          :::
998 gArgCysAlaValArgGlySerSerArgHisAspArgThrAlaSerThrA 1015

336 GCAGGCTTTGGACAAGGGCGAAGGGCTGC..................... 364
      |||              |||||||||
1015 rgArgProHisLysProProLysGlyCysAlaThrAspIleHisSerGly 1031

365 ......TGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGC 408
          |||      |||   :::|||:::   :::      |||:::::
1032 ArgTyrCysTrpProArgThrAlaSerSerArgAla......AlaSerGl 1046

409 TACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCC 458
    ::::||||||:::      |||                |||      |||
1046 yAlaSerAlaLysArgThrArgLeuArgArgArgSerCysProValArgS 1063

459 GAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCA 508
          :::|||   |||:::::::::   :::::::::|||:::::
1063 erProArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySer 1079

509 AAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAG 558
    :::::::::   ||||||   :::   :::      :::         ||
1080 SerSerArg...ArgProSerSerGlyArgProTrpSerValProIleAr 1095

559 GCCC.........TGCGCGCGGGCG.........AGGCAACCATCATCCT 590
    ||||         |||:::||||||              :::||||||||||
1095 gProSerSerIleCysGlyArgAlaValGlyLeuThrSerProSerSerP 1112

591 GCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCG.........TGT 631
       :::      :::   ||||||:::|||   |||          |||
1112 roLeuAsnArgProPheAlaArgArgSerAlaProAlaSerThrProCys 1128

632 GGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAA 681
      |||                                |||   |||:::::
1129 ArgArg............................HisAsnArgArgAr 1135

682 TTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCG....... 724
    ::::   :::   :::      :::      |||:::::::::
1135 gTyrGlySerArgArgProPheArgArgArgPheAlaCysSerTrpSerS 1152

724 ................................................. 724

1152 erGlnHisAspProAlaSerGlnAspProGlnArgGlyThrCysProLeu 1168

725 ...AACGCCTGCCCG......ACGGACAAGGCTTCG.............. 751
```

```
          ||||||||||||      :::   :::||||||
1169 ArgAsnAlaCysProGlyTrpAlaProArgAlaSerArgProHisLeuPr 1185

 752 .............TGT.........TGCACATCCGCCCCGTCCAAGGGGA 779
                     |||         |||          |||   ::::
1185 oLeuArgArgArgCysLysGluArgCysProPheArgCysSerProAlaS 1202

 780 ATTGAACGGCAACAAAGCCCACGATGCCG..............CCGTGT 814
     ::       ::::::::::||||||   |||              ||||||
1202 erProProThrAlaSerProThrTrpProAlaSerSerGlySerProCys 1218

 815 .................TCAACCGCAATACCGAATATTGGATACGCCGT 846
                         :::   |||      :::   :::::::::
1219 GlyAlaSerGlyAlaSerThrProAlaSerAlaSerTrpAlaHisSerAr 1235

 847 TTTCCGACGCAGTATCTGT 865
      ||||||   |||   |||
1235 gPheArgSerSerThrCys 1241
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAG33109

seq_documentation_block:

```
ID   AAG33109 standard; Protein; 337 AA.
XX
AC   AAG33109;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays protein fragment SEQ ID NO: 40068.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence; corn.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
```

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

```
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
           Quality:  127.50            Length:    271
             Ratio:  1.071               Gaps:     11
 Percent Similarity:  43.911   Percent Identity: 26.568

alignment_block:
 US-09-303-518D-571 x AAG33109   ..

 Align seg 1/1  to: AAG33109  from: 1  to: 337

     210 GCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAAC 259
         |||   ||||||:::|||      |||      ::::::         ::
      34 AlaThrGlyGlnAspArgArgProGlyAlaValArgLeuAlaAlaValVa 50

     260 TTGCCCCCGCGTTTTTCAAAAAACCGGA......AGACATCGAAACAATG 303
         :   |||   |||      :::   |||      |||:::|||::::::
      50 lProProValValAlaThrArgGlnGlyValArgArgArgArgArgArgA 67

     304 TTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGG 353
           |||:::   :::        :::|||                |||||
      67 rgGlnAspHisValLeuProProValArgArg.....GlnGly 79

     354 CGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCG 403
         ||||||      |||::::::      :::||||||   |||      ||
      80 ArgArgGlnGlnValArgArg......ArgArgGlnAlaArgArgLeuAr 94

     404 GACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGC......CATG 447
         |:::   |||      |||      :::      :::|||      |||:
      94 gAlaArgHisLeuGlyAlaProArgLeuGlyArgGluArgGlyGlyHisA 111

     448 TACAAGCCGCCGAAAATCAAAGCGAT...AGACAAAATCATGCAGGCGGG 494
         ::|||   |||:::   :::   |||   :::|||:::      ||||||
     111 laGlnGluAlaAspLeuArgArgAspProGluGlnAspValProGlyGly 127

     495 CAGGGT..................................... 500
         :::|||
     128 GluGlyHisGlyHisGlyGluProValProAspHisArgArgArgArgGl 144

     501 ................GCGCGGCAA..................... 509
                         |||||||||
     144 yLeuValProValGlnAlaArgGlnValProAspGluGluAlaValProG 161

     510 ................AGGCAAAACCGCGCCCACCGGCATACAAGGGGT 542
                        ||||||:::   :::|||||||:::   |||:::
     161 lyAlaAspValLeuHisArgGlnGlyGlyGlyHisArgGlnGluArgAla 177

     543 CAAACA..................................... 548

     178 AlaArgValProGluAspGlnAlaHisAspProProHisLeuHisAlaGl 194

     549 .AATCATCAAGGCCCTGCGCGCGGGCGA..................... 575
          :::   :::||||||||||||||||||
     194 yArgAspArgGlyProAlaArgGlyArgLeuArgArgHisProGlnValA 211

     576 ........GGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGA 617
                 |||:::|||:::      |||   :::      :::  |||
     211 rgGlyGluGlyArgHisArgLeuArgArgArgHisArgAlaAlaProGly 227

     618 AGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGA 667
         ||||||   |||:::      :::   ||||||::::::   :::|||
     228 ArgArgAlaArgAlaLeuProLeuHisArgGlnAlaAlaArgArgHisAr 244

     668 CACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 717
           :::|||::::::            ||||||      |||
     244 gGlnAlaGlyGlnLeu............ArgArgAlaLeuPro..... 254

     718 TGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCC 767
                         |||::::::|||              |||
     255 ..................ArgAlaGlnLeuProArgLeuLeuLeuPro 264

     768 CGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCA 817
         ||||||:::|||   :::|||   :::      |||   ||||||
     265 ArgProGlnGlyProArgArgLeuHisArgIleArgGlnArgArgArgAl 281

     818 ACCGCAATACCGA 830
          |||:::   |||
     281 aProArgArgArg 285

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAY20852

seq_documentation_block:
ID   AAY20852 standard; Protein; 201 AA.
XX
AC   AAY20852;
XX
DT   22-JUL-1999  (first entry)
XX
DE   Human neurofilament-H mutant protein fragment 11.
XX
KW   Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
KW   frameshift mutation; age-related disease; neurodegenerative disorder;
KW   Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
KW   Huntington's disease; multiple sclerosis; alcoholic liver disease;
KW   diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
KW   ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
KW   neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
KW   glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
KW   bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
KW   high mobility group protein-C; neuroendocrine specific protein A.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9845322-A2.
XX
PD   15-OCT-1998.
XX
PF   02-APR-1998;   98WO-IB00705.
XX
PR   10-APR-1997;   97US-0043163.
XX
PA   (UYUT-) RIJKSUNIV UTRECHT.
PA   (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
XX
PI   Burbach JPH,  Grosveld FG,  Van Leeuwen FW;
XX
DR   WPI; 1998-609901/51.
DR   N-PSDB; AAX75760.
XX
PT   Diagnosing disease by detecting frameshift mutations in RNA or
PT   corresponding protein mutations - used to diagnose cancer and
PT   neurological diseases, particularly Alzheimer's disease, and also
PT   for treatment and prevention with specific ribozymes or wild-type
PT   RNA
XX
PS   Disclosure; Figure 9; 258pp; English.
XX
CC   This invention describes a novel method for the diagnosis of a disease
CC   caused by, or associated with, an RNA molecule that has a frameshift
CC   mutation. The method is used to diagnose age-related diseases, especially
CC   cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC   disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC   multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC   and many others listed) or susceptibility to these disorders. The method
CC   allows a definitive diagnosis of Alzheimer's disease in living patients,
CC   at an early stage. It is based on the observation that disease may be
CC   caused by mutations in RNA rather than DNA. The invention describes the
CC   used of neuronal system RNA molecules, specifically proteins including
CC   beta-amyloid precursor protein (beta-APP), the microtubule associated
CC   proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
CC   associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
```

```
CC   neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
CC   protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC   2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC   protein-C (HMGP-C) and neuroendocrine specific protein A.
XX
SQ   Sequence   201 AA;


alignment_scores:
            Quality:  126.50            Length:    212
              Ratio:  1.360               Gaps:     10
 Percent Similarity:  43.868   Percent Identity: 28.774

alignment_block:
 US-09-303-518D-571 x AAY20852   ..

 Align seg 1/1  to: AAY20852  from: 1  to: 201

      69 GCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAA 118
         |||:::   ||||||      |||         :::|||   ||||||
          7 AlaGluArgProLeuArgArgValHisArgGlnGlyAlaAlaAlaGlyGl 23

     119 ACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGC 168
                       |||                 ||||||   |||:::|
          23 yAlaGlnProGlnProGly.............GlyArgArgArgValH 35

     169 ATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGT 218
         ||||||::      ||||||||||||   :::|||:::   :::
          35 isArgGlnGlyAlaAlaAlaGlyGlyAlaGlnProGlnProGly..... 49

     219 CAAAGCCGTTTTTGCGGA..............AACGGCAAAATGCGGTT 253
            :::|||   ||||||               :::|||   :::|||::
          50 ...GlyArgGlyCysGlyAlaAlaAlaAlaAlaGlyGlyProLeuArgTy 65

     254 TGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATG 303
         :|||   :::   |||:::
          65 rGlyArgAlaValArgAla............................. 71

     304 TTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGG 353
                 |||   |||   :::   |||:::|||:::   |||:::|||
          72 .....ArgGlyProArgAspAlaArgArgGlyAlaAlaProGlyArgGly 86

     354 CGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCG 403
            |||:::|||:::
          87 AlaArgSerAlaThr................................. 91

     404 GACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAG 453
                      |||:::|||::::::|||         |||
          92 .............ProGlyAlaGlyAlaProAlaArgGlyHisArgAlaA 104

     454 CCGCCGAAAATCAAAGCGATAGACAAAATCATGCA...GGCGGGCAGGGT 500
           |||                |||         |||   ||||||:::|||
         104 rgAlaProAlaProArgArgArgGlyProAlaAlaArgGlyGlyArgGly 120

     501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
         :::      |||   :::      :::         ||||||      ::
         121 GlyGlyProArgAlaGlyAlaLeuArgAlaGlyGlyArgGly......Gl 135

     551 TCATCAAGGCCCTGCGCGC............................... 569
         :   :::||||||||:::
         135 yAlaArgGlyProAlaGluGluGlyAlaGlyAlaAlaGlyGlyValArgL 152

     570 ..................GGGCGAGGCAACCATCATCCTGCCCGACCA 599
                           |||||||||:::   :::   |||||||||
         152 euProAlaAlaProProProGlyArgGlyGly...ArgAlaAlaArgPro 167

     600 CGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGC 635
            |||         |||   |||||||||:::|||
         168 AspPro.........GlyLeuArgArgArgAlaGly 176

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

ID   AAB59827 standard; Protein; 1592 AA.
XX
AC   AAB59827;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #4 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1592 AA;


alignment_scores:
            Quality:  126.50            Length:    297
              Ratio:  1.150               Gaps:     13
 Percent Similarity:  37.037   Percent Identity: 25.253

alignment_block:
 US-09-303-518D-571/rev x AAB59827   ..

 Align seg 1/1  to: AAB59827  from: 1  to: 1592

     814 ACACGGCGGCATCGTGGGCTTTGTTGCCGTTCAATTCCCCTTGGACGGGG 765
         |||||||||   |||:::   |||||||||:::             |||||
     670 ThrArgArgLysArgAsnCysCysCysArgAlaAlaArgArgPheArgGl 686

     764 C.................GGATGTGCAACACGAAGCCTTGTCCGTCGGG 733
         |                  :::||||||   |||         ||||||
     686 yProLeuArgAlaAlaProAlaCysAlaArgArgPheAlaTyrArgArgL 703

     732 CAGGCGTTCGCAGCAGAAAAACAGGGTTTTCACGCCTTTGACGTGTGCCA 683
           |||   |||      :::   |||   ||||||
     703 euGlyProArgCysThrSerArgGlyArgSerArgCys............ 715

     682 ATTTTGCCGCCAGTGTCATGGTGTATGCAGGTTTGCCGAAAAAATCCGCC 633
                       :::   |||   :::      |||   :::   ||
     716 .........SerProAspArgCysCysArgTrpSerArgCysSerSerPr 729
```

```
     632 CACACGCCGCCGCCTTCCTGCGGAGAAGGGACGTGGTCGGGCAGGATGAT 583
         |:::  ||||||  |||  ::::::     |||     |||
     729 oSerProArgArgCysProProSerSerProAlaGlyAlaProGlyAlaT 746

     582 GGTTG.....CCTCGCCCGCGCGCAGGGCCTTGA 554
                      |||     |||||||||
     746 hrCysSerArgArgProPheSerArgSerArgAspSerAlaGlyPro... 761

     553 TGATTTGTTTGACCCCTTGTATGCCGGTGGGCGCGGTTTTGCCTTTGCCG 504
                     |||  |||||||||     |||||
     762 .....ArgAlaAlaArgPheArgArgCysAr 770

     503 CGCACCCTGCCCGCCTGCATGATTTTGTCTATCGCTTTGATTTTCGGCGG 454
         |     |||                                  |||
     770 gAspAlaCysGluArg.....ArgAlaA 778

     453 CTTGT..... 449
           |||
     778 rgCysProGlyProArgSerAlaProSerIleArgArgGlySerArgAsp 794

     448 .....ACATGGCGGTCAGGTGGAACGGAAGCTGCTGGCTGATGTAG 408
               :::  ||||||  |||:::     :::|||     ::
     795 ArgSerArgAlaSerArgSerArgGlySerProLeuCysGlyAlaThrAl 811

     407 CGTCCGCCCAAATCGTAGCTGCCGATGTGCGGCGTGATGAACAGCAGCCC 358
         ::::        |||     ||||||:::           :::|||
     811 aThrSerCysProArgArgArgArgCysSer.....IleGlyAlaS 825

     357 TTCGCCCTTGTCCAAAGCCTGCTGCACGTGTTCCCAGCCGTGTACCGCTT 308
               ||||||  |||  :::|||        |||  |||:::
     825 erSerGlyCysProHisProProValArg.....ArgSerProVal 838

     307 TGAACATTG.....TTTCGATGTCTTCCGGTTTTTTGAAAAAC 270
           :::             ||||||          :::
     839 AsnSerSerLysArgAlaHisArgArgCysThrAlaArgArgGlyArgPh 855

     269 GCGGGGGCAAGTTCCA.....AACCGCATTTTGCCGTTTC 235
          ||||||  :::             :::|||        :::|
     855 eArgGlyProThrSerArgAspThrGlyArgArgArgCysTrpArgTrpP 872

     234 CGCAAAAACGGCTTTGACCGTCTGCGTGTCGG..... 203
         ||:::          |||     ||||||
     872 roArg.....ProArgArgCysArgCysSerArgArgTrpGly 884

     202 .....GGTTCAAACCCGCCTGCCGCATATTGGCGACGATGCGC 165
                 :::|||:::  |||  |||  |||||||||  ::
     885 ArgProLeuTrpAlaSerGlyCysProArgAlaArgTrpArgArgGlySe 901

     164 GCGCGGTCTTCCTTTAAAAGGTAAAACGCCAGATGTCCGAGCCGGTTTCC 115
         :                       :::     :::|||
     901 r.....AsnTrpSerSerGlyArgS 908

     114 CAGCGTGTGCAGACAGGAAAGCGACAGCAGGGAGAGGCATT 74
           ::::::|||:::  |||  |||:::||||||  :::
     908 erSerAlaAlaSerProLysArgThrCysGlyArgArgVal 921
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51580

seq_documentation_block:

```
ID   AAU51580 standard; Protein; 627 AA.
XX
AC   AAU51580;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #12476.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59551.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 12775; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   627 AA;
```

alignment_scores:

```
           Quality:  126.00              Length:    352
             Ratio:  0.900                 Gaps:     21
 Percent Similarity:  39.773   Percent Identity: 25.568
```

alignment_block:

US-09-303-518D-571 x AAU51580 ..

Align seg 1/1 to: AAU51580 from: 1 to: 627

```
      42 CGCCATGCACATCCT.....GTTGACCGCCCTGCTCAAAT 76
         |||||||:::  |||           :::||||||  :::
     312 ArgHisGlyAspProSerGlyProArgThrAlaAspArgHisAspAspHi 328

      77 GCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTC 126
          |||                    |||:::||||||:::|||
     328 sPro.....ArgAlaArgAlaGlyArgPro.... 336

     127 GGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC 176
                    |||     :::  |||  :::  :::|||:::
     337 .....AspAspLeuLeuGlnArgHisArgLeuGlnLeuArgArgHis 350

     177 CAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCG 226
```

```
351 Gln.........SerGlyHisArgProCysHisLeuArgArgGlnSerAr 364

227 TTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTC 276
364 gArgGluGly.................GlyThrCysGlyArgIleArgG 375

277 AAAAA..................................................... 281
375 lyArgLysValHisLeuGlyGluProAlaAlaArgProLeuProGlyHis 391

282 ACCGGAAGACAT.........CGAAACAATGTTCAAAGCGGTACACGGCT 322
392 ArgGlyArgHisHisGlyGlyArgGlnArgTyrLeuSerGlyAspPro.. 407

323 GGGAACACGTGCAGCAGGCTTTGGACAAGGG..................... 353
408 .GlyCysLeuAlaThrProTyrArgSerGlyLeuProGlySerGlyThrV 424

354 .................CGAAGGGCTGCTGTTCATCACGCCGCACATCGG 386
424 alLeuArgIleAspProGluGluHisArgLeuArgHisArgGlyHisArg 440

387 CAGCTA.....................CGATTTGGGCGGACGCTACATCA 415
441 ArgAlaAspProArgArgCysSerProArgLeuArgArgProLeuHisGl 457

416 GCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATC 465
457 n......AlaValLeuProAlaValArgHisProArg............. 467

466 AAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAA 515
468 ............................ArgGlyAlaArgTyrGlnAla 474

516 AACCGC............GCCCACCGGCATACAAGGGGTCAAACAAATCA 553
475 LeuArgTrpProGluAlaAlaHis.................LeuAspCy 485

554 TCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTC 603
485 sGlnGlyProAlaGlnGlyArgProAspProArgSerArgArgGlyHis. 501

604 CCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACC 653
502 ..LeuSerProGlyHisGlnValArgAlaValGlyAlaGlyArgSerGly 517

654 TGCATACAC.................CATGACACTGGCGGCAAAATTGG 685
518 IleProHisGlyArgProHisGlnProHisHisArgProProProLeuAs 534

686 CACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCC 735
534 pAspLeuGlnArgArgProAsp..............ArgHisProGlyA 546

736 GACGGACAAGGCTTCGTGTTGCACATCCGCCCCGT.............. 770
546 rgArgSerArgArgArgAspArgHisProGlyArgAlaGlyArgLeuArg 562

771 .......................CCAAGGGGAATTGAA........... 785
563 ArgTyrLeuArgArgValValGlyProAlaGlyIleGlnLeuGluGlyAr 579

786 ......................CGGCAACAA..............AGCC 798
579 gProGlnGluAlaGlyAlaIleArgHisGlnThrValMetArgThrAlaP 596

799 CACGATGCCGCCGTGTTCAACCGCAATAC......CGAATATTGGATACG 842
596 roThrMetAlaProThrGlnProArgHisGluArgArgIleValProThr 612

843 CCGTTT 848
```

```
613 ArgPhe 614

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04993

seq_documentation_block:
ID   AAY04993 standard; Protein; 637 AA.
XX
AC   AAY04993;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 49B.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34244.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 49B; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   637 AA;


alignment_scores:
            Quality:  125.00              Length:    266
              Ratio:  1.068                 Gaps:     14
 Percent Similarity:  43.985   Percent Identity: 26.692

alignment_block:
 US-09-303-518D-571 x AAY04993   ..

 Align seg 1/1  to: AAY04993  from: 1  to: 637

     153 GGAAGACCGCGC.................GCGCATCGTCGCCAATATGC 184
     276 GlyArgProArgGlyCysGlnAlaGluProAlaArgArgValAlaArg** 292

     185 GGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTT.... 230
     292 *AlaGlyGlyPheArgProArg......ProGlyHisSerArgAlaGlyA 307

     231 ................................TGCGGAAACGGCAAA 245
     307 laProSerGlyVal***ThrValLeuProValGlyIleGlyThrGlyVal 323
```

```
     246 ATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCG 295
             ||||||      |||||||||
     324 AlaGlyPheGly......ProArgVal..................... 330

     296 AAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTG 345
                                 |||   |||   :::|||   :
     331 ..........................GlyAspArgGlnThrGlyGlyA 338

     346 GACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGA 395
         ::|||   ||||||              ||||||||||||:::
     338 laGlnProArgArgIle............AlaAlaHisArgArg...... 348

     396 TTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCA 445
                     |||:::|||         :::|||   :::
     349 ..............HisArgProArgArgProAlaProTrpAsnValAs 360

     446 TGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCA...... 488
          ::::::||||||   :::   ||||||   ::::::   |||
     360 pLeuArgAlaAlaProArgProSerAspAlaAspSerAlaAlaSerArgC 377

     489 ..............GGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCC 524
                       :::   :::|||   |||:::   ::::::   :::
     377 ysAspGlyTrpArgSerGluHisGlyHisArgGluLeuSerGlyPheGly 393

     525 CACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGG.. 572
         :::|||:::      |||   :::   |||         ||||||:::
     394 GluArgTyrLeuSerGlyIleSerValHis.........AlaArgAsnLe 407

     573 ................................CGAGGCAACCATCATC 588
                                                :::      |
     407 uValValGlnLeuLeuLeuLysAlaThrProMetProValHisAlaTrpP 424

     589 CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGA 638
         ||:::   |||:::   :::::::::   |||:::   :::::::::
     424 roSerTrpProGlu***TrpAla***TyrArgLysSerAsnLeuAla... 439

     639 TTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCAC 688
                 :::   |||   |||                   |||
     440 .........Thr***ArgIleThrHis.................GlyIl 447

     689 ACGTCAAAGGCGTGAAAAC......CCTGTTTTTCTGCTGCGAACGCCTG 732
          |||:::||||||::::::      |||      :::||||||      :
     447 eArgHisArgArgSerSerArgGluProGlnValMetLeuArgGlySerS 464

     733 CCCGACGGACA.........AGGCTTCGTGTTGCACATCCGCCCCGTCCA 773
         :::::||||||                ||||:::|||   |||||||||
     464 erLysArgThrAlaSerSerSerThrGlyValThrHisArgProArgPro 480

     774 AGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCG 821
         |||:::               :::|||      |||   |||   |||
     481 ArgSerGlnProSerThrValAsnProThrArgArgTrpValValPro 496

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29221

seq_documentation_block:
ID   AAY29221 standard; Protein; 447 AA.
XX
AC   AAY29221;
XX
DT   25-OCT-1999  (first entry)
XX
DE   Amino acid sequence of a virulence factor encoded by ORF31266.
XX
KW   Human pathogen; virulence polypeptide; virulence factor;
KW   pathogenic infection; Pseudomonas aeruginosa infection.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO9927129-A1.
XX
PD   03-JUN-1999.
XX
PF   25-NOV-1998;   98WO-US25247.
XX
PR   25-NOV-1997;   97US-0066517.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Ausubel F,  Cao H,  Drenkard E,  Goodman HM,  Mahajan-Miklos S;
PI   Rahme LG,  Tan M,  Tsongalis J;
XX
DR   WPI; 1999-357851/30.
XX
PT   Virulence factors useful in developing disease treatments
XX
PS   Disclosure; Fig 3; 228pp; English.
XX
CC   The present sequence represents a Pseudomonas aeruginosa polypeptide
CC   sequence. P. aeruginosa is an opportunistic human pathogen present in
CC   soil water and plants. The specification describes virulence polypeptides
CC   and nucleic acid sequence encoding such polypeptides. These sequences
CC   can be used to identify a compound which is capable of decreasing the
CC   expression of a pathogenic virulence factor. Compounds that inhibit
CC   the expression or activity of virulence factor polypeptides can be
CC   used to treat pathogenic infections, especially where the infection
CC   is a P. aeruginosa infection.
CC   note: the sequences given in the specification were poorly legible, and
CC   in some instances assumptions were made as to the identity of the
CC   residue; it is therefore possible that the sequence given below is
CC   not entirely correct.
XX
SQ   Sequence   447 AA;


alignment_scores:
           Quality:  123.00              Length:    327
             Ratio:  0.904                Gaps:     17
 Percent Similarity:  41.590   Percent Identity: 24.771

alignment_block:
 US-09-303-518D-571 x AAY29221   ..

 Align seg 1/1  to: AAY29221  from: 1  to: 447

      18 CAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTT......GA 61
         :::|||||||   |||      :::|||||||||   |||:::      ::
      19 ArgAlaValArgProValProHisArgHisAla...ProAlaGlyArgGl 34

      62 CCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACG 111
         :||||||   :::   |||:::                         |||
      34 nArgProArgSerGlyProVal.......................HisP 43

     112 CTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCG 161
           :::   |||:::         |||
      43 roAlaAlaProSerThrAspProGly.....................Gly 52

     162 CGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGC 211
         :::   ||||||   :::                   :::|||   |||
      53 HisAspHisArgValArg.................GlnProGlyHisAr 63

     212 AGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTT 261
          |||      ::::::                            |||   |
      63 gAspProGlyAlaGln.........................GlyArgC 71

     262 GCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGC 311
         ||   |||:::   ::::::||||||||   |||:::   :::   :::
      71 ysLeuArgLeuProHisGlnThrGlyArgProArgSerLeuAlaGlyAla 87

     312 GGT.................ACACGGCTGGGAACACGTGCAGCAGGCTT 343
         |||                 ::::::   |||:::|||:::::::::
      88 GlyGlyAsnArgProThrLeuAlaGlnProGlySerArgGlySerAlaGl 104

     344 TGGACAA......GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGC 387
          ||||||      :::|||||||:::|||:::              :::
     104 yGlyGlnProProAlaArgArgValAlaAlaAspAlaArgProAlaGlnP 121
```

```
     388 AGCTACGATTT..............GGGCGGACGCTACATCAG..... 416
             |||                 |||      |||||||||
     121 roAspArgGlnAlaGlyAlaGlnProGlyAlaGlyLeuHisGlnTrpArg 137

     417 .................................CCAGCAGCTTCCGT 430
                                          |||   |||:::::
     138 ValArgGlnArgGlnGlyThrGlyGlyAlaProAspProArgAlaGlyAl 154

     431 TCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAA 480
         :             |||:::|||:::|||   :::   |||      :
     154 aThrTyrArgAlaAlaValArgAlaGlyGluLeuArgArgAspSerLeuA 171

     481 ATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGG 530
         ::      |||   :::      |||   ::::::::::   |||
     171 rgAlaAspGlyLysArgValLeuArgProGlnGluArgGlnLeuHisTrp 187

     531 CATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAA 580
         :::   |||   |||:::         ||||||:::   ||||||   ::
     188 ArgTyrArgArgGlnAla.........GlyProValProGlyArgGlnAr 201

     581 CCATCATCCTGCC..............CGACCACGTCCCTTCTCCGCAG 615
         :   ||||||:::              ||||||          ||||
     201 gTrpHisProValProArgArgSerArgArgProAlaAspGlyHisAlaG 218

     616 GA..................................AGGCGGCGGCGT 629
         ||                                 |||   ||||||
     218 lyGlnThrAlaProGlyAspProGlyLysGlyArgAlaArgGlyArgArg 234

     630 GTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAA 679
            :::|||      |||:::||||||         :::::::::
     235 ProAlaGlyGlyArgArgArgThrCysAla......SerSerAlaProPr 249

     680 AATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACG. 728
             |||         ::::::  |||           ||||||
     249 oThrArgThrSerProProLysSerAlaProGlyAlaSerAlaArgThrS 266

     729 .................................CCTGCCCGACGGACA 743
                                          |||   ||||||
     266 erThrThrAlaSerThrSerSerSerCysAlaTyrThrAlaAlaArgThr 282

     744 AGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACA 793
                                 |||||||||
     283 ........................ProArgGly.............. 285

     794 AAGCCCACGATGCCGCCGTGTTCAACCGCAA 824
             |||   :::|||||:::::||||||
     286 .HisProAlaAlaArgArgThrHisProGln 295

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146

seq_documentation_block:
ID   ABG05146 standard; Protein; 440 AA.
XX
AC   ABG05146;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #5137.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS69333.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 35505; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   440 AA;


alignment_scores:
           Quality:  122.00              Length:    336
             Ratio:  0.797                 Gaps:     16
 Percent Similarity:  45.536   Percent Identity: 25.298

alignment_block:
 US-09-303-518D-571 x ABG05146   ..

 Align seg 1/1  to: ABG05146  from: 1  to: 440

      27 TCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAAT 76
         :::|||:::|||:::|||   ::::::   :::      |||:::
       7 AlaProLeuAlaHisArgTrp***GlnLysLeuGlyLeuProGlyCysTh 23

      77 GCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTC 126
          |||      :::   :::||||||||||   :::            ||||
      23 rPro...ThrValCysSerPheLeuSerProGlnProProProArgAlaA 39

     127 GGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC 176
         ||:::|||::::::         :::|||   |||||| |||
      39 rgSerSerSerAlaThrLeuSerSerGlyAlaProArg.Arg........ 52

     177 CAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACG.......... 216
                      :::   :::|||:::   ||||||
      53 ................AlaAlaThrProGlnProGlnThrThrProProS 64

     217 .....GTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTT 261
               |||:::         |||      :::      |||   :::
      64 erSerValGlnProCysAlaAlaThrProCysAlaArgGlyGlyArgPro 80

     262 GCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGC 311
         |||||||||            |||    |||      :::   :::   |
      81 AlaProAlaGlyValThrTrpPro.ThrThrArgProSerMetAla***A 97
```

```
312 GGTACACG...GCTGGGAACACGTGCAGCAGGCTTTGG...ACAAGGGCG 355
    ||   :::   |||::::::|||:::   |||:::         |||:::
 97 rgThrSer***AlaSerThrThrAlaProArgMetAlaProProArgSer 113

356 AAGGGCTGCTGTTCATCACGCCGC.....ACATCGGCAGCTAC 393
       :::|||:::   |||:::|||            ||||||   ||
114 HisAlaCysAlaArgSerHisArgProGluThrAlaArgSerAlaArgTh 130

394 GATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGC 443
    |         :::|||   :::      :::|||   ||||||   ||||
130 rAlaProArgSerAlaIleThrArgArgAlaPheThrSerThrArgProP 147

444 CATGTACAAGCCGC..................................... 457
    ||   |||:::|||
147 roProThrThrArgThrValAlaSerSerGlyThrHisThrSerGlyLeu 163

458 .....................................CGAAAATC 465
                                         |||:::||
164 SerProThrAlaSerArgLeuAlaArgCysArgAlaProGlyArgSerSe 180

466 AAAGCGATAGACAAAATCATGCAGG.........CGGGCAGGGTGCGCGG 506
    |         |||   :::|||||||         |||      |||   |
180 rThrIleIleThr***ThrCysArgSerProThrArgLeuCysCysProA 197

507 CAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCA... 553
    ||:::   :::|||   |||||||||   :::   |||::::::
197 laGln...ArgProLeuProProAlaSerSerProSerSerSerArgThr 212

554 ..............TCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATC 588
                  :::|||   |||:::|||   |||      |||||
213 SerArgSerValTrpThrArgArgCysThrArgLeuArgTrpThrSerSe 229

589 CTGCCCGACCACGTCCCTT...CTCCGCAGGAAGGCGGCGGCGTGTGGGC 635
    |   |||   :::   :::   ||||||      :::::::::   |||
229 rArgProProSerTrpMetAlaLeuArgThrValGlyThrSerThrGlyP 246

636 GGATTTTTTCGGCAAACCTGCAT......ACACCATGACACTGGCGGCAA 679
           :::   :::|||:::                  ||||||:::
246 roThrAla***ArgSerLeuArgArgCysGlnAlaSerThrTrpArgSer 262

680 AATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGC 729
    :::      ||||||         :::   |||        |||   ::
263 ArgProSerThrSerAlaProProSerTrpCysAlaArgTrpAlaAlaTh 279

730 CTGCCCGACGGACAAGGCT.............................. 748
    :   |||      :::|||
279 r***ProLeuProSerAlaCysGlnArgLysTrpSerMetLeuTrpArgT 296

749 ......................TCGTGTTGCACATCCGCCCCGTCCAAG 775
                          |||:::|||   :::|||||||||
296 hrGlyThrAlaArgValSerThrSerAlaCysGlyAlaAlaProSerThr 312

776 GGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAAT 825
    ::::::   ::::::::::::|||:::   |||        :::   ::
313 SerArg...SerThrSerArgProSerThrProMetLeuArgAlaProVa 328

826 ACCG 829
    :|||
328 lPro 329
```